Iowa State University
Digital Repository

Retrospective Theses and Dissertations | Iowa State University Capstones, Theses and Dissertations

1992

On the multivariate components of variance problem

Sellem Remadi
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Remadi, Sellem, "On the multivariate components of variance problem " (1992). *Retrospective Theses and Dissertations*. 10343.
https://lib.dr.iastate.edu/rtd/10343

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9234846

On the multivariate components of variance problem

Remadi, Sellem, Ph.D.

Iowa State University, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

On the multivariate components of variance problem

by

Sellem Remadi

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1992

Copyright © Sellem Remadi, 1992. All rights reserved.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr Yasuo Amemiya for his guidance and constructive criticism during the period of preparation of this dissertation.

I extend my gratitude to Dr Koehler and Dr Sacks who taught me everything I know about applied statistics.

I also thank Dr Harville for his comments and suggestions during the five years I spent at I. S. U.

Finally, I would like to thank my parents and my wife Amel for their unconditional support

1. GENERAL INTRODUCTION

1.1 Problem and Literature Review

The univariate mixed linear models have been discussed often in the statistical literature and have been used widely in applied sciences. In many applications, it is assumed that the data vector follows the general mixed effect linear model

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Z}\mathbf{b} + \mathbf{e}, \tag{1.1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observed values, the $n \times k$ $\mathbf{X}$ and $n \times q$ $\mathbf{Z}$ are given explanatory or incidence matrices, γ is the $k \times 1$ vector of fixed effects, $\mathbf{b}$ is the $q \times 1$ vector of random effects, and $\mathbf{e}$ is the $n \times 1$ vector of random errors. This univariate variance component problem has been discussed extensively. Considering applications in animal science, Henderson (1953) proposed an estimation procedure obtained by equating the mean squares to their expected values to solve for the variance components. Herbach (1959), Thompson (1962), and Patterson and Thompson (1971, 1974) proposed procedures for the maximum likelihood and the restricted (residual) maximum likelihood estimation of the variance components. Miller (1977) discussed asymptotic properties of the maximum likelihood estimator. Harville (1977) and Searle et al. (1992) provide summaries and reviews for the univariate problem.

Unlike the univariate case, development for the multivariate covariance component models has been limited. The standard multivariate version of model (1.1) for a p-dimensional response each on N observations is

$$\mathbf{Y} = \mathbf{X\Gamma} + \mathbf{ZB} + \mathbf{E}, \tag{1.2}$$

where $\mathbf{Y}$ is the $N \times p$ matrix of observations, the $N \times k$ $\mathbf{X}$ and $N \times n$ $\mathbf{Z}$ are given explanatory or incidence matrices, $\mathbf{\Gamma}$ is the $k \times p$ matrix of fixed effects, $\mathbf{B}$ is the $n \times p$ matrix of random effects, and the $N \times p$ $\mathbf{E}$ is the matrix of random errors. This model assumes that each of the random effects in the model has the same number of variables as the response variable. The random effects given as $1 \times p$ rows of $\mathbf{B}$ are classified into t effects with the $i-th$ effect having n_i levels with common covariance matrix Σ_{ii}. The levels of each effect are assumed to be uncorrelated with each other, the levels of the other effect, and the residual or error effects in E which are assumed to have common covariance matrix $\Sigma_{t+1,t+1}$. The matrices $\Sigma_{11}, \ldots, \Sigma_{t+1,t+1}$ are called the covariance components. This model and special cases are discussed in, e.g., Thompson (1973), Amemiya (1985), Meyer (1985), and Anderson and Amemiya (1991).

Taking $p = 1$ in model (1.2), we obtain the univariate model (1.1). On the other hand, applying the vec operator to model equation (1.2), the multivariate model (1.2) can be regarded as a univariate model. However, because of the complexity of the covariance structure and of the restrictions on the parameter space, the univariate representation of the multivariate model does not allow the use of standard univariate results. Furthermore, a class of statistical problems associated with the multivariate model (1.2) does not arise in the univarite case. One such problem is that of possibly reduced rank of a covariance component. For the univariate or multivariate mixed models, a question of interest is whether a particular random effect is significant.

The hypothesis of no $i-th$ random effect is that the $i-th$ variance or covariance component is zero. However, for the multivariate case, a random effect can exist but be restricted to a space of dimension less than p. Thus, in a sense, the existence of the $i-th$ random effect consists of different structures specified by rank $\Sigma_{ii} = 1, 2, \ldots, p$. Hence, the question of rank should be asked at the model building stage in fitting model (1.2). With some knowledge of the rank, one can generally develop more efficient estimators of the covariance components. It turns out that properties of an estimator of a covariance component are affected by the ranks of this and other covariance components. Therefore, recognition of the rank problem is important for estimation problem as well. In this dissertation, testing and estimation problems for the multivariate model are discussed, not as applications or extensions of the standard univariate results, but as special topics associated with the rank problem.

One important issue in developing asymptotic results for variance covariance component problems is the choice of index over which the limit is taken. For example, the one way random effect model has two indices, the numbers of groups and replicates. In practice, the general model can appear in a variety of sampling configurations. To develop useful asymptotic procedures, various assumptions about sampling configuration have to be considered. In this dissertation, asymptotic properties of statistical procedures are derived under a wide range of assumptions. Our goal is to develop procedures useful for most practical situations without worrying about which asymptotic case a particular sample corresponds to.

1.2 Explanation of Dissertation Format

This dissertation consists of three papers. In the first paper, we discuss the limiting distributions of the roots of a certain determinantal equation. The original motivation for considering such limiting distributions was to develop necessary results for the multivariate covariance component problem. In the multivariate problem, the roots of a certain determinantal equation play important roles in testing for the rank and in estimation of covariance components. It turns out that such necessary results are not available in the literature. It was decided to investigate the general limiting distribution results which contain cases required for the covariance component problem as special cases. What makes our general results different from the existing standard results is the difference in the rates of convergence for the two random matrices as well as for different roots.

In the second paper, we address the probem of testing the rank of the covariance matrix of a random effect. Based on the results of the first paper, several test procedures for testing the rank of the between-group covariance matrix in the unbalanced multivariate mixed effect model with one-way random structure are proposed and their asymptotic null distributions are derived. Exact (non asymptotic) tests are derived for a certain simple special case. Test procedures which can be used for a large range of practical situations are proposed. Such test procedures are then extended to more complicated models. Also, we report the results of a simulation study for comparing these test procedures.

The third paper is devoted to the estimation problems in the balanced multivariate one-way random effect model. The restricted maximum likelihood estimators of covariance components derived under a rank condition on the between-group component are considered. Under the assumption that such a rank condition may be incorrect, and under various assumptions regarding the rate of increase of the numbers of groups and replicates, asymptotic properties of the estimators are derived. Such discussion leads to recommendation of an approximate inference procedure which can be used in various practical situations. The results of a simulation study are reported.

The papers are followed by a general summary and the references cited in the general introduction are in the reference section that follows the general summary.

PAPER I.

LIMITING DISTRIBUTION OF ROOTS WITH DIFFERENTIAL RATES OF CONVERGENCE AND ITS APPLICATIONS

ABSTRACT

Some general results on the limiting distribution of the roots of the determinantal equation $|\mathbf{A}(n) - \lambda \mathbf{B}(n)| = 0$ are given, and their applications to the multivariate covariance component problems are discussed.

1. INTRODUCTION

Let $n = 1, 2, \ldots,$ be some sequence over which the limit is taken, and let $\mathbf{A}(n)$ and $\mathbf{B}(n)$ be $p \times p$ random symmetric matrices indexed by n. Assume that $\mathbf{B}(n)$ is positive definite with probability one for each n. Then, with probability one, the determinantal equation in λ

$$|\mathbf{A}(n) - \lambda\mathbf{B}(n)| = 0 \tag{1.1}$$

has p ordered real roots $\hat{\lambda}_1(n) \geq \hat{\lambda}_2(n) \geq \ldots \geq \hat{\lambda}_p(n)$. We are concerned with the limiting distribution of properly normalized $\hat{\lambda}_i(n)$ when normalized $\mathbf{A}(n)$ and $\mathbf{B}(n)$ have a joint limiting distribution as $n \to \infty$.

Anderson (1984b) contains the standard results on the asymptotic distribution of the roots when $\mathbf{A}(n)$ and $\mathbf{B}(n)$ are matrices arising in certain multivariate analyses. The limiting distribution of the roots of (1.1) dates back to Hsu (1941) and Anderson (1951) who considered a noncentral Wishart matrix $\mathbf{A}(n)$ with degrees of freedom not depending on n and a central Wishart matrix $\mathbf{B}(n)$ with degrees of freedom increasing with n. Anderson (1989a) considered the case where $\mathbf{A}(n)$ and $\mathbf{B}(n)$ are the mean square matrices in multivariate components of variance. Amemiya (1990) gave a general characterization of the limiting distribution of $\hat{\lambda}_i(n)$. He assumed that for some diagonal matrix $\mathbf{D}(n) = \text{diag}\{d_1(n), \ldots, d_p(n)\}$

$$\sqrt{n}[\mathbf{A}(n) - \mathbf{D}(n),\ \mathbf{B}(n) - \mathbf{I}_p] \xrightarrow{L} [\mathbf{U}, \mathbf{V}], \tag{1.2}$$

and considered the limiting distribution of

$$\sqrt{n}[\hat{\lambda}_i(n) - d_i(n)],\ i = 1, 2, \ldots, p. \tag{1.3}$$

Note that if we let $\mathbf{B}(n) = \mathbf{I}_p$ for all n then the $\hat{\lambda}_i(n)$'s are the eigenvalues of $\mathbf{A}(n)$. Thus, the case of eigenvalues of a single matrix can be considered a special case of the general form (1.1). Deriving a result for the form (1.1) with two matrices $\mathbf{A}(n)$ and $\mathbf{B}(n)$ from a result for the eigenvalues of a single matrix is difficult in most applications, because assumptions are usually given and verified in terms of the joint limiting distribution of $\mathbf{A}(n)$ and $\mathbf{B}(n)$, and not of $\mathbf{B}^{-1/2}(n)\mathbf{A}(n)\mathbf{B}^{-1/2}(n)$. The eigenvalue case with $\mathbf{B}(n) = \mathbf{I}_p$ has attracted substantial interest in the literature. See, e.g., Anderson (1963), Waternaux (1976), Davis (1977), Fujikoshi (1980), Fang and Krishnaiah (1982), and Eaton and Tyler (1991). In these papers, $\mathbf{A}(n)$ satisfies the part of (1.2) corresponding to $\mathbf{A}(n)$, and the limiting distribution or expansion for the quantities of the form (1.3) is derived.

Our initial motivation for investigating the limiting distribution of the $\hat{\lambda}_i(n)$'s under a nonstandard condition arose from our study of multivariate components of variance problems. In its simplest form, we assume a $p \times 1$ vector $\mathbf{Y}_{jl}$ satisfies

$$\mathbf{Y}_{jl} = \mu + \alpha_j + \epsilon_{jl},\ j = 1, 2, \ldots, J,\ l = 1, 2, \ldots, L, \tag{1.4}$$

where μ is a $p \times 1$ fixed vector, and α_j and ϵ_{jl} are $p \times 1$ independent random vectors with zero mean. The between-group covariance matrix is $\mathbf{\Phi} = Var\{\alpha_j\}$ of rank $r \leq p$. The within-group covariance matrix $\mathbf{\Sigma} = Var\{\epsilon_{jl}\}$ is assumed to be positive definite. Let the between-group and within-group mean square matrices be denoted by

$$\mathbf{M}_b = \frac{L}{J-1}\sum_{j=1}^{J}(\bar{\mathbf{Y}}_{j.}-\bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{j.}-\bar{\mathbf{Y}}_{..})',$$

$$\mathbf{M}_w = \frac{1}{J(L-1)}\sum_{j=1}^{J}\sum_{l=1}^{L}(\mathbf{Y}_{jl}-\bar{\mathbf{Y}}_{j.})(\mathbf{Y}_{jl}-\bar{\mathbf{Y}}_{j.})',$$

where $\bar{\mathbf{Y}}_{j.} = \Sigma_{l=1}^{L}\mathbf{Y}_{jl}/L$ and $\bar{\mathbf{Y}}_{..} = \Sigma_{j=1}^{J}\bar{\mathbf{Y}}_{j.}/J$. For this multivarite one-way balanced covariance components model, standard estimators of $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ as well as test statistics for the rank r and fit of the model are functions of the roots of

$$\left|\mathbf{M}_b - \lambda\mathbf{M}_w\right| = 0. \tag{1.5}$$

See, e.g., Anderson (1984a, 1989b), Amemiya and Fuller (1984), Schott and Saw (1984), Anderson et al. (1986), and Anderson and Amemiya (1991). Let $\mathbf{T}$ be a $p \times p$ nonsigular matrix such that $\mathbf{T\Phi T}' = \mathrm{diag}\{\phi_1,\ldots,\phi_p\}$, $\phi_1 \geq \ldots \geq \phi_p$, and $\mathbf{T\Sigma T}' = \mathbf{I}_p$. Then, the roots of (1.5) are the roots $\hat{\lambda}_i$'s of (1.1) with

$$\begin{aligned}\mathbf{A}(n) &= \mathbf{TM}_b\mathbf{T}', \\ \mathbf{B}(n) &= \mathbf{TM}_w\mathbf{T}'.\end{aligned} \tag{1.6}$$

Thus, the limiting distribution of the $\hat{\lambda}_i$'s is relevant. To consider limiting behavior, let n be the index over which the limit is taken, and assume that in general the number of groups $J = J(n)$ and the number of replicates $L = L(n)$ are functions of n. For the case where L is a constant (free of n) and $J(n) = n \to \infty$, Anderson (1989a) derived the limiting distribution of the $\hat{\lambda}_i$'s. This case can also be covered by Amemiya's (1990) general result, because $\mathbf{A}(n)$ and $\mathbf{B}(n)$ in (1.6) satisfy (1.2) with nontrivial $\mathbf{U}$ and $\mathbf{V}$. Thus, the p quantities in (1.3) have a nondegenerate joint limiting distribution.

A question arises when the number of replicates $L = L(n)$ increases with n. Asymptotics for this problem when $L(n)$ increases and $J(n)$ either increases or is a constant have not been discussed previously, but they are important in developing asymptotic approximate inference procedures for a variety of practical situations including multivariate variance component models more complicated than the balanced one-way model. To see why the case with increasing $L(n)$ is not covered by existing results, assume that $r = \text{rank}(\mathbf{\Phi}) < p$, *i.e.*, $\phi_r > \phi_{r+1} = \ldots = \phi_p = 0$, and let

$$\begin{aligned} \mathbf{A}(n) &= \begin{pmatrix} \mathbf{A}_{11}(n) & \mathbf{A}_{12}(n) \\ \mathbf{A}21(n) & \mathbf{A}_{22}(n) \end{pmatrix}, \\ \mathbf{D}_1(n) &= L(n)\ \text{diag}\{\phi_1, \ldots, \phi_r\} + \mathbf{I}_r, \end{aligned} \tag{1.7}$$

where $\mathbf{A}_{11}(n)$ is $r \times r$. Then, it follows that

$$J^{1/2}[L^{-1}(n)\{\mathbf{A}_{11}(n) - \mathbf{D}_1(n)\};\ L^{-1/2}(n)\mathbf{A}_{21}(n);\ \mathbf{A}_{22}(n) - \mathbf{I}_{p-r}], \tag{1.8}$$

has a nondegenerate limiting distribution as $n \to \infty$, but that $J^{1/2}(n)[\{\mathbf{A}_{11}(n) - \mathbf{D}_1(n)\}; \mathbf{A}_{21}(n)]$ tends to infinity. Hence, the parts of $\mathbf{A}(n)$ have different rates of convergence, and the $\mathbf{A}(n)$ part of (1.2) does not hold for a nondegenerate $\mathbf{U}$ (even with redefined n). Furthermore, a proper normalizing constant for $\mathbf{B}(n)$ is $J^{1/2}(n)L^{1/2}(n)$ which differs from the normalizing constant for any part of $\mathbf{A}(n)$. Thus Amemiya's (1990) result does not apply if a nondegenerate limiting distribution of all p roots $\{\hat{\lambda}_i,\ i = 1, \ldots, p\}$ is sought. As we shall show in section 3, quantities of the form (1.3) do not have a joint nondegenerate limiting distribution because the r larger roots and the $p - r$ smaller roots have different rates of convergence.

Anderson (1951) considered one particular situation where the roots have different rates of convergence. This involves a noncentral Wishart matrix with constant degrees of freedom and a reduced rank noncentral Wishart matrix tending to infinity. The source of the differential rates is a fixed (not random) noncentral structure, which is distinct from the random effects variance component problem. In section 2, we present results on the limiting distribution of the roots with differential rates for a broad class of problems. The class is characterized by a general structure of differential rates which is different from Anderson's. Our structure corresponds to that appearing in the one-way and more complicated multivariate variance component problems as well as in multivariate structural relationship (errors-in-variables) problems. In our structure, the rates of convergence for $\mathbf{A}(n)$ and $\mathbf{B}(n)$ are generally different, and the four parts of $\mathbf{A}(n)$ have particular types of differential rates. Applications of our results to the one-way variance component model are discussed in section 3. Derivations of the results are given in section 4. For the results in section 2, the basic method of proof follows that given by Anderson (1951) and Amemiya (1990). Arguments similar to those of Anderson (1989b) and Anderson and Amemiya (1991) are utilized in section 3.

2. LIMITING DISTRIBUTION RESULTS

We present results on the limiting distribution of the roots $\hat{\lambda}_i(n)$'s of (1.1) without assuming specific distributional form of $\mathbf{A}(n)$ and $\mathbf{B}(n)$ except that $\mathbf{A}(n)$ and $\mathbf{B}(n)$ have a joint limiting distribution when standardized by a general form of normalizing sequences. To define our general structure, let a_n, b_n, and c_n be sequences of constants indexed by $n = 1, 2, \ldots$. Assume that each a_n, b_n, and c_n is either a constant (free of n) or a sequence tending to infinity as $n \to \infty$. We present the limiting distribution assumption on $\mathbf{A}(n)$ and $\mathbf{B}(n)$ after a canonical transformation so that $\mathbf{A}(n)$ and $\mathbf{B}(n)$ are centered around a diagonal $\mathbf{D}(n)$ and $\mathbf{I}_p$, respectively. The diagonal elements of $\mathbf{D}(n)$ depend on n in general, and play roles of the "true roots" around which the observed roots $\hat{\lambda}_i(n)$ are centered. To present our assumption on the differential rates of convergence, let

$$\begin{aligned}
\mathbf{U}(n) &= a_n \begin{pmatrix} c_n^{-1}\mathbf{I}_r & 0 \\ 0 & \mathbf{I}_{p-r} \end{pmatrix} [\mathbf{A}(n) - \mathbf{D}(n)] \begin{pmatrix} c_n^{-1}\mathbf{I}_r & 0 \\ 0 & \mathbf{I}_{p-r} \end{pmatrix} \\
&= a_n[\mathbf{A}^*(n) - \mathbf{D}^*(n)], \qquad (2.1) \\
\mathbf{V}(n) &= b_n[\mathbf{B}(n) - \mathbf{I}_p].
\end{aligned}$$

Using the notation in (1.7), the normalizing sequence in $\mathbf{U}(n)$ is $a_n c_n^{-2}$ for $\mathbf{A}_{11}(n)$, $a_n c_n^{-1}$ for $\mathbf{A}_{21}(n)$, and a_n for $\mathbf{A}_{22}(n)$. Note that our interest is in the p roots $\hat{\lambda}_i(n)$'s of (1.1) which are in general entirely dfferent from the p roots of $\left|\mathbf{A}^*(n) - \gamma\mathbf{B}(n)\right| = 0$.

We assume:

(i) as $n \to \infty, [\mathbf{U}(n), \mathbf{V}(n)] \xrightarrow{L} [\mathbf{U}, \mathbf{V}]$.

To represent a general multiplicity structure of the "true roots", we assume that the first r diagonal elements of $\mathbf{D}(n)$ are grouped into s distinct values with multiplicities $q_{1j}, j = 1, 2, \ldots, s$, and the last $p-r$ diagonal elements into t values with multiplicities $q_{2j}, j = 1, 2, \ldots, t$. That is,

$$\begin{aligned} \mathbf{D}(n) &= \text{block diag}\{\mathbf{D}_1(n), \mathbf{D}_2(n)\} \qquad (2.2) \\ &= \text{block diag}\{d_{11}(n)\mathbf{I}_{q_{11}}, \ldots, d_{1s}(n)\mathbf{I}_{q_{1s}},\ d_{21}(n)\mathbf{I}_{q_{21}}, \ldots, d_{2t}(n)\mathbf{I}_{q_{2t}}\}. \end{aligned}$$

Note that $\mathbf{D}^*(n)$ has the same structure with $d^*_{kj}(n)$ replacing $d_{kj}(n)$ where $d^*_{1j}(n) = c^{-2}(n)d_{1j}(n)$ and $d^*_{2j}(n) = d_{2j}(n)$. We assume:

(ii) as $n \to \infty, d^*_{kj}(n) \to d^*_{kj}$ for all k and j, where $d^*_{11} > \ldots > d^*_{1s} > 0$ and $d^*_{21} > \ldots > d^*_{2t}$.

Note that there is no restriction in (ii) for the relative sizes of d^*_{1j} and d^*_{2j}. This is because the r larger roots and $p - r$ smaller roots are distinguished easily by differential rates of convergence. There may appear to be some indeterminacy in choices of a_n and c_n in verifying assumption (i). We first choose c_n to have proper limits of the "true roots" in (ii), and then choose a_n in (i) for such a c_n so that $\mathbf{U}(n)$ has a proper limit. To center the observed roots $\hat{\lambda}_i(n)'s$ around the "true roots" $d_{kj}(n)$'s, define for $k = 1, 2$ and $j = 1, \ldots, s$ (for $k = 1$) or t (for $k = 2$)

$$\begin{aligned} p_{kj} &= \sum_{m=1}^{j} q_{km}, \\ \mathbf{g}_{kj}(n) &= [\hat{\lambda}_{p_{k,j-1}+1}(n) - d_{kj}(n), \ldots, \hat{\lambda}_{p_{kj}}(n) - d_{kj}(n)], \\ \mathbf{h}_1(n) &= [\mathbf{g}_{11}(n), \ldots, \mathbf{g}_{1s}(n)], \qquad (2.3) \end{aligned}$$

$$\mathbf{h}_2(n) = [\mathbf{g}_{21}(n), \ldots, \mathbf{g}_{2t}(n)],$$

where $p_{10} = 0$, $p_{1s} = p_{20} = r$ and $p_{2t} = p$. Note that $\mathbf{g}_{kj}$ is the $1 \times q_{kj}$ vector of the roots centered around $d_{kj}(n)$, and that $\mathbf{h}_1$ and $\mathbf{h}_2$ are the $1 \times r$ and $1 \times (p-r)$, respectively, corresponding to the larger r roots and smaller $(p-r)$ roots. We divide the limiting matrices $\mathbf{U}$ and $\mathbf{V}$ in (i) in the form

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix}, \ \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \tag{2.4}$$

where $\mathbf{U}_{11}$ and $\mathbf{V}_{11}$ are $r \times r$, and $\mathbf{U}_{22}$ and $\mathbf{V}_{22}$ are $(p-r) \times (p-r)$.

The sequences a_n, b_n and c_n represent rates of convergence typically appearing in statistical applications, and are assumed to be monotonically nondecreasing sequences of positive numbers. To avoid trivial cases, assume:

(iii) as $n \to \infty$, $b_n \to \infty$ and $c_n \to \infty$.

We distinguish two cases depending on whether a_n tends to infinity or is a constant. Both cases appear in our application to the covariance component problem. We first consider the case with $a_n \to \infty$ and assume:

(iv) as $n \to \infty$, $a_n \to \infty$ and $\frac{a_n}{b_n} \to \gamma$, where γ is a nonnegative number or ∞.

For this case with $a_n \to \infty$ under (ii) and (iv), let

$$\begin{aligned} \mathbf{D}_1^* &= \text{block diag}\{d_{11}^*\mathbf{I}_{q_{11}}, \ldots, d_{1s}^*\mathbf{I}_{q_{1s}}\}, \\ \mathbf{D}_2^* &= \text{block diag}\{d_{21}^*\mathbf{I}_{q_{21}}, \ldots, d_{2t}^*\mathbf{I}_{q_{2t}}\}, \end{aligned} \tag{2.5}$$

and for $k = 1, 2$ let

$$\begin{aligned} \mathbf{G}_{kk} &= \mathbf{U}_{kk}, && \text{if } \gamma = 0, \\ &= \mathbf{U}_{kk} - \gamma \mathbf{D}_k^* \mathbf{V}_{kk}, && \text{if } 0 < \gamma \leq 1, \end{aligned} \tag{2.6}$$

$$= \gamma^{-1}\mathbf{U}_{kk} - \mathbf{D}_k^*\mathbf{V}_{kk}, \qquad \text{if } 1 < \gamma < \infty,$$
$$= -\mathbf{D}_k^*\mathbf{V}_{kk}, \qquad \text{if } \gamma = \infty,$$

where $\mathbf{U}_{kk}$ and $\mathbf{V}_{kk}$ are defined in (2.4). We also write

$$\mathbf{G}_{11} = \begin{pmatrix} \mathbf{G}_{11}^{11} & . & . & . & \mathbf{G}_{11}^{1s} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \mathbf{G}_{11}^{s1} & . & . & . & \mathbf{G}_{11}^{ss} \end{pmatrix}, \quad \mathbf{G}_{22} = \begin{pmatrix} \mathbf{G}_{22}^{11} & . & . & . & \mathbf{G}_{22}^{1t} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \mathbf{G}_{22}^{t1} & . & . & . & \mathbf{G}_{22}^{tt} \end{pmatrix}, \tag{2.7}$$

where $\mathbf{G}_{kk}^{jl}$ is $q_{kj} \times q_{kl}$.

Theorem 1. Under (i)-(iv), as $n \to \infty$,

$$min\{a_n, b_n\}[c_n^{-2}\mathbf{h}_1(n),\ \mathbf{h}_2(n)] \xrightarrow{L} (\mathbf{z}_1, \mathbf{z}_2),$$

where $\mathbf{z}_1 = (\mathbf{z}_{11}, \ldots, \mathbf{z}_{1s})$, $\mathbf{z}_2 = (\mathbf{z}_{21}, \ldots, \mathbf{z}_{2t})$, and $\mathbf{z}_{kj}$ is the $1 \times q_{kj}$ vector consisting of the q_{kj} ordered eigenvalues of $\mathbf{G}_{kk}^{jj}$,

Note that the distribution of $(\mathbf{z}_1, \mathbf{z}_2)$ in Theorem 1 is the nondegenerate joint limiting distribution of all p roots $\hat{\lambda}_i(n)$'s of (1.1) after proper standardization. The r largest roots and $(p-r)$ smallest roots have different rates of convergence in the sense that the normalizing constants are $min\{a_n, b_n\}c_n^{-2}$ and $min\{a_n, b_n\}$ respectively. If we relax (iii) and assume $b_n \to \infty$ but $c_n = 1$, then the parts of $\mathbf{A}(n)$ do not have differential rates. For this standard case, Theorem 1 is still valid with an additional assumption $d_{1s}^* > d_{21}^*$ in (ii). The form of $\mathbf{G}_{kk}$ in (2.6) may appear counter-intuitive, especially if $\gamma = \infty$. But, $\mathbf{G}_{kk}$ is the leading term in some expansion, and so depends only on the slowest converging term. An intuitive appeal of $\mathbf{G}_{kk}$ may be obtained by considering a special case with $p = 1$.

We next consider the case where a_n is a constant, i.e., a normalizing constant for $\mathbf{A}_{22}(n)$ is one. Without loss of generality, we assume:

(v) $a_n = 1$ for all n.

For this case, let $\mathbf{F}_{11} = \mathbf{U}_{11}$ and $\mathbf{F}_{22} = \mathbf{U}_{22} - \mathbf{U}_{21}(\mathbf{U}_{11} + \mathbf{D}_1^*)^{-1}\mathbf{U}_{12}$, where the quantities are defined in (2.4) and (2.5). Let $\mathbf{F}_{kk}^{jl}$ denote the $q_{kj} \times q_{kl}$ submatrix of $\mathbf{F}_{kk}$, $k = 1, 2$, as in (2.7).

Theorem 2. Under (i)-(iii), (v) as $n \to \infty$

$$[c_n^{-2}\mathbf{h}_1(n),\ \mathbf{h}_2(n)] \xrightarrow{L} (\mathbf{w}_1, \mathbf{w}_2),$$

where $\mathbf{w}_1 = (\mathbf{w}_{11}, \ldots, \mathbf{w}_{1s})$, $\mathbf{w}_2 = (\mathbf{w}_{21}, \ldots, \mathbf{w}_{2t})$, $\mathbf{w}_{kj}$ is the $1 \times q_{kj}$ vector consisting of the q_{kj} ordered eigenvalues of $\mathbf{F}_{kk}^{jj}$.

Note that the nondegenerate limiting distribution given by $(\mathbf{w}_1, \mathbf{w}_2)$ does not depend on $\mathbf{V}$ because the convergence of $\mathbf{B}(n)$ is faster than $\mathbf{A}(n)$. Using (ii), we can rephrase the result of Theorem 2 as

$$\{c_n^{-2}[\hat{\lambda}_1(n), \ldots, \hat{\lambda}_r(n)],\ [\hat{\lambda}_{r+1}(n), \ldots, \hat{\lambda}_p(n)]\} \xrightarrow{L} (\mathbf{y}_1, \mathbf{y}_2), \tag{2.8}$$

where $\mathbf{y}_1 = (\mathbf{y}_{11}, \ldots, \mathbf{y}_{1s}), \mathbf{y}_2 = (\mathbf{y}_{21}, \ldots, \mathbf{y}_{2t})$, $\mathbf{y}_{kj}$ is the $1 \times q_{kj}$ vector consisting of the q_{kj} ordered eigenvalues of $\mathbf{E}_{kk}^{jj} = \mathbf{F}_{kk}^{jj} + d_{kj}^*\mathbf{I}_{q_{kj}}$. If we write the part of (i) and (ii) corresponding to $\mathbf{A}(n)$ as

$$\mathbf{A}^*(n) \xrightarrow{L} \mathbf{U} + \mathbf{D}^* = \mathbf{U}^* = \begin{pmatrix} \mathbf{U}_{11}^* & \mathbf{U}_{12}^* \\ \mathbf{U}_{21}^* & \mathbf{U}_{22}^* \end{pmatrix}, \tag{2.9}$$

with $r \times r$ $\mathbf{U}^*_{11}$, then $\mathbf{E}^{jj}_{11}$ is a submatrix of $\mathbf{U}^*_{11}$ and $\mathbf{E}^{jj}_{22}$ is a submatrix of $\mathbf{U}^*_{22} - \mathbf{U}^*_{21}\mathbf{U}^{*-1}_{11}\mathbf{U}^*_{12}$. This form of theorem 2 may be easier to deal with in application.

3. APPLICATION TO THE COVARIANCE COMPONENT PROBLEM

Consider the multivariate covariance component model given in (1.4). For simplicity, assume that the within-group error $\epsilon_{jl} \sim N_p(\mathbf{0}, \Sigma)$. For the between-group effect, we only assume that α_j's are identically distributed with covariance matrix $\mathbf{\Phi}$ and finite fourth order moments. Statistical inferences on the rank of $\mathbf{\Phi}$ is of interest. The between-group effect may exist in a space of dimension less than p . Also the knowledge on rank $\mathbf{\Phi}$ can be used in devoloping efficient inference procedures. If α_j's are also normally distributed, then the likelihood ratio test statistic for the hypothesis that rank $\mathbf{\Phi} \leq r$ is

$$\begin{aligned} T &= -2log(LR) = \sum_{i=r+1}^{p} I(\hat{\lambda}_i > 1) f(\hat{\lambda}_i), \qquad (3.1) \\ f(\lambda) &= (JL-1)log\frac{(J-1)\lambda + J(L-1)}{JL-1} - (J-1)log\lambda, \end{aligned}$$

where $\hat{\lambda}_i$'s are the roots of (1.5) , $I(.)$ is the indicator function, J is the number of groups, and L is the number of replicates. See Anderson et al. (1986). We suppress the theoretical index n in this section. To obtain an approximate cut-off point, we can consider three limiting cases: (I) $J \to \infty$ and L fixed, (II) $J \to \infty$ and $L \to \infty$, and (III) J fixed and $L \to \infty$. Thus, we might end up with three different asymptotic cut-off points. In most practical situations, it may not be clear which of the three cut-off points should be used for a particular sample. Hence, it would be nice to develop a test procedure which can be justified under various cases, i.e., which gives an approximately correct type I error under a wide range of practical situations. The limiting distribution of T for case (I) when the true rank of $\mathbf{\Phi}$ is r was derived by Anderson (1989b), and was tabulated in Amemiya et al. (1990). The distribution is not chi-squared. Cases (II) and (III) have not been discussed previously in the literature. We note that T is a function of the $(p-r)$ smallest roots $\hat{\lambda}_i$'s of (1.5) . The limiting distribution of such roots follows immediately from Theorem 1 for cases (I) and (II), and from Theorem 2 for case (III).

Assuming that the true rank of $\mathbf{\Phi}$ is r, assumptions (i) and (ii) hold for $\mathbf{A}(n)$ and $\mathbf{B}(n)$ of (1.6) with $a_n = J^{1/2}, b_n = J^{1/2}L^{1/2}, c_n = L^{1/2}$, $t = 1$, $q_{21} = p - r$, and $d_{21}^* = 1$. Note that $min\{a_n, b_n\} = J^{1/2}$, and $\gamma = lim\ b_n^{-1}a_n$ in (iv) is $L^{-1/2}$ for case (I) and 0 for case (II). Thus, for these two cases, by Theorem 1 and the ensuring discussion, $J^{1/2}(\hat{\lambda}_{r+1} - 1, \ldots, \hat{\lambda}_p - 1)$ converges to the $(p-r)$ ordered eigenvalues of $\mathbf{G}_{22}$ in (2.6) which is $\mathbf{U}_{22} - L^{-1/2}\mathbf{V}_{22}$ for case (I) and $\mathbf{U}_{22}$ for case(II). As can be seen from the transformation (1.6), the distribution of $\mathbf{U}_{22}$ does not depend on the distribution of α_j's. For case (II), the distribution of the symmetric $\mathbf{G}_{22} = \mathbf{U}_{22}$ consists of independent normal random variables with mean 0 and variance 2 for diagonal and 1 for off-diagonal elements. For case (I), the distribution of $\mathbf{G}_{22} = \mathbf{U}_{22} - L^{-1/2}\mathbf{V}_{22}$ is $L^{1/2}(L-1)^{-1/2}$ times that of $\mathbf{G}_{22}$ for case (II). But, applying these limiting distributions for $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ to expand the test statistic T in (3.1), we find that for cases (I) and (II) the limiting distributions of T are identical. The common limiting distribution is that given by Anderson (1989b) for case (I) with normal α_j's and tabulated in Amemiya et al. (1990). Note that the result here was shown for any distribution of α_j's as well as for case (II). Thus, the table in Amemiya et al. (1990) has a wider range of use than originally intended. The use of T is justified when J is large. However, case (III) is slightly different in that we are trying to make inferences on the rank of the between-group covariance matrix without the number J of the between-group effect vectors α_j tending to infinity. For this case, Theorem 2 applies, and by (2.8) and (2.9) $(\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p)$ coverges to the $(p-r)$ ordered eigenvalues of $\mathbf{E}_{22} = \mathbf{U}^*_{22} - \mathbf{U}^*_{21}\mathbf{U}^{*-1}_{11}\mathbf{U}^*_{12}$. The conditional distribution of $(J-1)$ times $\mathbf{U}^*$ in (2.9) given (finitely many) α_j's is the noncentral Wishart distribution with covariance matrix diag$\{\mathbf{0}, \mathbf{I}_{p-r}\}$, degrees of freedom $(J-1)$, and the noncentrality matrix block diag$\{\mathbf{\Psi}, \mathbf{0}\}$ for some $r \times r$ $\mathbf{\Psi}$. Thus, the conditional and unconditional distribution of $(J-1)\mathbf{E}_{22}$ is $W_{p-r}(\mathbf{I}_{p-r},\ J-r-1)$. Hence, even for finite J, the limiting distribution of $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ is common for any distribution of α_j's with finite fourth moments. It follows that the limiting distribution of the test statistic T for case (III) is the distribution of

$$\sum_{i=1}^{p-r} (J-1)I(\delta_i > 1)(\delta_i - 1 - log\delta_i), \tag{3.2}$$

where $(J-1)\delta_i$'s are the eigenvalues of a $W_{p-r}(\mathbf{I}_{p-r},\ J-1-r)$ matrix. This distribution differs from the limiting distribution for cases (I) and (II), and has not been tabulated. A practitioner may not feel comfortable in using the test statistic T and the cut-off point in Amemiya et al. (1990) when J is not large, because such a cut-off point is unjustified by asymptotic approximation for case (III).

To develop a test procedure which can be justified in cases including (III), let

$$S = (J-1) \sum_{i=r+1}^{p} \hat{\lambda}_i.$$

For case (III), S converges to $(J-1)\,\mathrm{tr}(\mathbf{E}_{22})$ which has the χ^2_d distribution with $d=(p-r)(J-r-1)$ for any distribution of α_j's with finite fourth moments. For case (II), $d^{-1/2}(S-d)$ converges to $\mathrm{tr}(\mathbf{G}_{22})$, i.e., $N(0,2)$ which is also the limiting distribution of $d^{-1/2}(\chi^2_d - d)$. In this sense, the use of S and a χ^2_d cut-off point provides an approximately correct size for cases (II) and (III). For case (I), S converges to $N(0, 2L(L-1)^{-1})$, and the χ^2_d cut-off point may make the test too liberal for small L. For practical situations where L is large (but J may not be large), the use of S and χ^2_d cut-off points is recommended.

4. DERIVATIONS

This section provides derivations of Theorems 1 and 2 in section 3, and some details for section 4. In derivations of Theorems 1 and 2, we assume, for simplicity, $p = 2$ and $r = 1$, i.e., $\mathbf{A}(n)$ and $\mathbf{B}(n)$ are 2×2 and 2 roots $\hat{\lambda}_1(n)$ and $\hat{\lambda}_2(n)$ have differential rates. We simply use $d_1(n)$, $d_2(n)$, $d_1^*(n)$, $d_2^*(n) = d_2(n), d_1^*$, and d_2^* to denote $d_{1j}(n)$, $d_{2j}(n)$, $d_{1j}^*(n)$, $d_{2j}^*(n)$, d_{1j}^*, and d_{2j}^*. Note that $s = t = 1$. The extension of our derivations to general p and general multiplicity can be done easily by revising the derivations in Amemiya (1990) to incorporate the arguments presented here. Our derivations here follow the basic approach in Anderson (1951) and Amemiya (1990) using Rubin's theorem reported and derived in Anderson (1963). Because of Rubin's theorem, it suffices to show the limiting distribution results in Theorems 1 and 2 as the limiting result for a nonrandom sequence when $n \to \infty$ and $[\mathbf{U}(n), \mathbf{V}(n)] \to [\mathbf{U}, \mathbf{V}]$ treated as a nonrandom sequence. This step is derived here for the case $p = 2$. See the references mentioned above for more on the use of Rubin's theorem.

Proof of Theorem 1. Let $m_n = min\{a_n, b_n\}$, and let

$$\begin{aligned} \eta_1(n) = m_n c_n^{-2} h_1(n) &= m_n c_n^{-2}[\hat{\lambda}_1(n) - d_1(n)], \\ \eta_2(n) = m_n h_2(n) &= m_n[\hat{\lambda}_2(n) - d_2(n)], \end{aligned}$$

where $h_1(n)$ and $h_2(n)$ are defined in (2.3) for our case here with $s = t = 1$. We show $[\eta_1(n), \eta_2(n)] \to (z_1, z_2)$ in Theorem 1 as a nonrandom sequence with assumption (i) holds as a nonrandom sequence. We first note that the ordered roots of

$$\left|c_n^{-2}\mathbf{A}(n) - \nu\mathbf{B}(n)\right| = 0 \tag{4.1}$$

are $c_n^{-2}\hat{\lambda}_1(n)$ and $c_n^{-2}\hat{\lambda}_2(n)$. By (i) and (ii), $c_n^{-2}\mathbf{A}(n) \to \operatorname{diag}\{d_1^*, 0\}$ and $\mathbf{B}(n) \to \mathbf{I}_2$. Thus, by the continuity of the roots,

$$\begin{aligned} c_n^{-2}\hat{\lambda}_1(n) &\to d_1^*(n), \\ c_n^{-2}\hat{\lambda}_2(n) &\to 0. \end{aligned} \tag{4.2}$$

To consider the limit of $\eta_1(n)$, note that the ordered two roots of a determinantal equation in ϕ

$$\left|c_n^{-2}\mathbf{A}(n) - [d_1^*(n) + m_n^{-1}\phi]\mathbf{B}(n)\right| = 0 \tag{4.3}$$

with $d_1^*(n) = c_n^{-2}d_1(n)$ are $\hat{\phi}_1 = \eta_1(n)$ and $\hat{\phi}_2 = m_n[c_n^{-2}\hat{\lambda}_2(n) - d_1^*(n)]$. Note that $m_n \to \infty$ by (iii) and (iv), and that $d_1^*(n) \to d_1^* > 0$ by (ii). Thus, by (4.2) ,

$$\hat{\phi}_2 \to -\infty. \tag{4.4}$$

We multiply $\left|\operatorname{diag}\{m_n^{1/2}, 1\}\right|$ to (4.3) from both right and left, and write the result in a single determinant to obtain an equivalent equation

$$\begin{vmatrix} P_{11}(n) & P_{21}(n) \\ P_{21}(n) & P_{22}(n) \end{vmatrix} = 0, \tag{4.5}$$

where by writing $\mathbf{A}(n)$ and $\mathbf{B}(n)$ in terms of $\mathbf{U}(n)$ and $\mathbf{V}(n)$ in (2.1)

$$\begin{aligned} P_{11}(n) &= m_n a_n^{-1}U_{11}(n) - [1 + b_n^{-1}V_{11}(n)]\phi - d_1^*(n)m_n b_n^{-1}V_{11}(n), \\ P_{21}(n) &= m_n^{1/2}a_n^{-1}c_n^{-1}U_{21}(n) - [d_1^*(n) + m_n^{-1}\phi]m_n^{1/2}b_n^{-1}V_{21}(n), \\ P_{22}(n) &= c_n^{-2}d_2^*(n) + a_n^{-2}c_n^{-1}U_{22}(n) - [d_1^*(n) + m_n^{-1}\phi][1 + b_n^{-1}V_{22}(n)]. \end{aligned}$$

As $n \to \infty$, as $[\mathbf{U}(n), \mathbf{V}(n)] \to (\mathbf{U}, \mathbf{V})$ treated as a nonrandom sequence, and as $d_k^*(n) \to d_k^*$

$$\begin{aligned} P_{11}(n) &\to G_{11} - \phi, \\ P_{21}(n) &\to 0, \\ P_{22}(n) &\to -d_1^* < 0, \end{aligned}$$

where G_{11} is defined in (2.6). Thus, the coefficients of the second degree polynomial in ϕ given by (4.5) converge to the coefficients in $|G_{11} - \phi| \left|d_1^*\right| = 0$. This fact, (4.4), and Lemma 1 of Amemiya (1990) imply that

$$\hat{\phi}_1 = \eta_1(n) \to G_{11}, \tag{4.6}$$

where G_{11} is the only root of $|G_{11} - \phi| = 0$ in this case of $p = 2$.

To consider the limit of $\eta_2(n)$, note that the ordered two roots of

$$\left|\mathbf{A}(n) - [d_2(n) + m_n^{-1}\psi]\mathbf{B}(n)\right| = 0 \tag{4.7}$$

are $\hat{\psi}_1 = m_n[\hat{\lambda}_1(n) - d_2(n)]$ and $\hat{\psi}_2 = \eta_2(n)$. By (4.2)

$$\hat{\psi}_2 \to \infty. \tag{4.8}$$

We multiply $\left|\mathrm{diag}\{c_n^{-1}, m_n^{1/2}\}\right|$ to (4.7) from both right and left, and write the resulting product as

$$\begin{vmatrix} Q_{11}(n) & Q_{12}(n) \\ Q_{21}(n) & Q_{22}(n) \end{vmatrix} = 0, \tag{4.9}$$

where in terms of $\mathbf{U}(n)$ and $\mathbf{V}(n)$ in (2.1)

$$\begin{aligned} Q_{11}(n) &= d_1^*(n) + a_n^{-1}U_{11}(n) - c_n^{-2}[d_2^*(n) + m_n^{-1}\psi][1 + b_n^{-1}V_{11}(n)], \\ Q_{21}(n) &= m_n^{1/2}a_n^{-1}U_{21}(n) - [d_2^*(n) + m_n^{-1}\psi]m_n^{1/2}c_n^{-1}b_n^{-1}V_{21}(n), \\ Q_{22}(n) &= m_n a_n^{-1}U_{22}(n) - m_n b_n^{-1}d_2^*(n)V_{22}(n) - \psi - b_n^{-1}V_{22}(n), \end{aligned}$$

As $n \to \infty$, as $[\mathbf{U}(n), \mathbf{V}(n)] \to (\mathbf{U}, \mathbf{V})$, and as $d_k^*(n) \to d_k^*$,

$$\begin{aligned} Q_{11}(n) &\to d_1^*, \\ Q_{21}(n) &\to 0, \\ Q_{22}(n) &\to G_{22} - \psi, \end{aligned}$$

where G_{22} is defined in (2.6). Thus, the limiting polynomial for (4.9) is $\left|d_1^*\right| |G_{22} - \psi| = 0$. Hence, by Lemma 1 of Amemiya (1990) and (4.8),

$$\hat{\psi}_2 = \eta_2(n) \to G_{22}. \tag{4.10}$$

Since (4.6) and (4.10) give the joint limit for $[\eta_1(n), \eta_2(n)]$ as a nonrandom sequence, the result follows from Rubin's theorem by noting $z_1 = G_{11}$ and $z_2 = G_{22}$ for $p = 2$.

Proof of Theorem 2. We derive the limit of $[c_n^{-2} h_1(n), h_2(n)]$ treated as a nonrandom sequence when $a_n = 1$ for all n. By (i), (ii), and (v), $c_n^{-1}\mathbf{A}(n) \to \operatorname{diag}\{U_{11} + d_1^*,\ 0\}$ and $\mathbf{B}(n) \to I_2$. Thus,the ordered roots of (4.1) satisfy

$$\begin{aligned} c_n^{-2}\hat{\lambda}_1(n) &\to U_{11} + d_1^*, \\ c_n^{-2}\hat{\lambda}_2(n) &\to 0. \end{aligned} \tag{4.11}$$

Hence, it follows that

$$c_n^{-2} h_1(n) = c_n^{-2}\hat{\lambda}_1(n) - d_1^* \to U_{11} = F_{11}. \tag{4.12}$$

To consider the limit of $h_2(n)$, note that the ordered roots of

$$\left|\mathbf{A}(n) - [d_2^*(n) + \psi]\mathbf{B}(n)\right| = 0 \tag{4.13}$$

are $\hat{\psi}_1 = \hat{\lambda}_1(n) - d_2^*(n)$ and $\hat{\psi}_2 = h_2(n)$. By (4.11),

$$\hat{\psi}_1 \to \infty. \tag{4.14}$$

We multiply $\left|\mathrm{diag}\{c_n^{-1}, 1\}\right|$ to (4.13) from both left and right, and write the resulting equation as

$$\begin{vmatrix} Q_{11}(n) & Q_{12}(n) \\ Q_{21}(n) & Q_{22}(n) \end{vmatrix} = 0, \tag{4.15}$$

where in terms of $[\mathbf{U}(n), \mathbf{V}(n)]$ in (2.1)

$$\begin{aligned}
Q_{11}(n) &= d_1^*(n) + U_{11}(n) - c_n^{-2}[d_2^*(n) + \psi][1 + b_n^{-1}V_{11}(n)], \\
Q_{21}(n) &= U_{21}(n) - c_n^{-1}[d_2^*(n) + \psi]b_n^{-1}V_{21}(n), \\
Q_{22}(n) &= U_{22}(n) - b_n^{-1}d_2^*(n)V_{22}(n) - \psi[1 + b_n^{-1}V_{22}(n)].
\end{aligned}$$

As $n \to \infty$, as $[\mathbf{U}(n), \mathbf{V}(n)] \to [\mathbf{U}, \mathbf{V}]$, and as $d_k^*(n) \to d_k^*$,

$$\begin{aligned}
Q_{11}(n) &\to U_{11} + d_1^*, \\
Q_{21}(n) &\to U_{21}, \\
Q_{22}(n) &\to U_{22} - \psi,
\end{aligned}$$

Thus, the coefficients of the polynomial in ψ given by (4.15) converge to the coefficients of the polynomial

$$\begin{vmatrix} U_{11} + d_1^* & U_{21} \\ U_{21} & U_{22} - \psi \end{vmatrix} = \left|U_{11} + d_1^*\right| \left|U_{22} - U_{21}^2(U_{11} + d_1^*)^{-1} - \psi\right| = 0.$$

Hence, by (4.14) and Lemma 1 of Amemiya (1990),

$$h_2(n) \to U_{22} - U_{21}^2(U_{11} + d_1^*)^{-1} = F_{22}. \tag{4.16}$$

The result follows from (4.12) and Rubin's theorem.

Some Details for Section 3. Since the rank $\mathbf{\Phi}$ = rank $\mathrm{Var}\{\alpha_j\} = r$, $\mathbf{T}$ defined in (1.6) satisfies $\mathbf{T\Phi T'}$ = block diag$\{\mathbf{\Phi}_0, \mathbf{0}\}$ and $\mathbf{T\Sigma T'} = \mathbf{I}_p$, where $\mathbf{\Phi}_0 = \mathrm{diag}\{\phi_1, \ldots, \phi_r\}$ with $\phi_r > 0$. The roots $\hat{\lambda}_i(n)$'s of (1.5) are the roots of $|\mathbf{A}(n) - \lambda\mathbf{B}(n)| = 0$ for $\mathbf{A}(n)$ and $\mathbf{B}(n)$ defined in (1.6). Thus, without loss of generality, we assume that $\mathbf{\Phi} = \mathrm{Var}\{\alpha_j\}$ = block diag$\{\mathbf{\Phi}_0, \mathbf{0}\}$ and $\mathbf{\Sigma} = \mathrm{Var}\{\epsilon_{jl}\} = \mathbf{I}_p$, and that with probability one $\alpha_j = \left(\beta_j', \mathbf{0}\right)'$ for $r \times 1$ β_j with $\mathrm{Var}\{\beta_j\} = \mathbf{\Phi}_0 = \mathrm{diag}\{\phi_1, \ldots, \phi_r\}$. Let

$$\begin{aligned}
\delta_j &= (\delta_{1j}', \delta_{2j}')' = L^{1/2}\bar{\epsilon}_{j.} = L^{-1/2}\sum_{l=1}^{L}\bar{\epsilon}_{jl},\\
\bar{\delta} &= (\bar{\delta}_1', \bar{\delta}_2')' = J^{-1}\sum_{j=1}^{J}\delta_j,\\
\bar{\beta} &= J^{-1}\sum_{j=1}^{J}\beta_j.
\end{aligned}$$

Then, $\mathbf{A}(n)$ in (1.6) can be written

$$\mathbf{A}(n) = \begin{pmatrix} \mathbf{A}_{11}(n) & \mathbf{A}_{12}(n) \\ \mathbf{A}21(n) & \mathbf{A}_{22}(n) \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{A}_{11}(n) &= \frac{1}{J-1}\sum_{j=1}^{J}(L^{1/2}\beta_j - L^{1/2}\bar{\beta} + \delta_{1j} - \bar{\delta}_1)(L^{1/2}\beta_j - L^{1/2}\bar{\beta} + \delta_{1j} - \bar{\delta}_1)',\\
\mathbf{A}21(n) &= \frac{1}{J-1}\sum_{j=1}^{J}(\delta_{2j} - \bar{\delta}_2)(L^{1/2}\beta_j - L^{1/2}\bar{\beta} + \delta_{1j} - \bar{\delta}_1)',\\
\mathbf{A}_{22}(n) &= \frac{1}{J-1}\sum_{j=1}^{J}(\delta_{2j} - \bar{\delta}_2)(\delta_{2j} - \bar{\delta}_2)'.
\end{aligned}$$

Note that $\mathbf{B}(n)$ is independent of $\mathbf{A}(n)$ and $J(L-1)\mathbf{B}(n) \sim W_p(\mathbf{I}_p,\ J(L-1))$. Note also that the β_j's have fourth moments and $\delta_j \sim N_p(0,\ \mathbf{I}p)$. Thus, for case (I) with constant L and $J \to \infty$, $\mathbf{U}(n)$ in (2.1) with $a_n = J^{1/2}$ and $c_n = L^{1/2}$ converges to a matrix of normal random variables. For case (II) with $L \to \infty$ and $J \to \infty$,

$$\mathbf{U}(n) \;=\; J^{1/2}\begin{pmatrix} \frac{1}{J-1}\sum_{j=1}^{J}(\beta_j-\bar{\beta})(\beta_j-\bar{\beta})' - \mathbf{\Phi}_0 & \frac{1}{J-1}\sum_{j=1}^{J}(\beta_j-\bar{\beta})(\delta_{2j}-\bar{\delta}_2)' \\ \frac{1}{J-1}\sum_{j=1}^{J}(\delta_{2j}-\bar{\delta}_2)(\beta_j-\bar{\beta})' & \mathbf{A}_{22}(n)-\mathbf{I}_{p-r} \end{pmatrix} + O_p(L^{-1/2}) \tag{4.17}$$

converges to a matrix of normal random variables. For both cases (I) and (II), $\mathbf{U}_{22}$ which is the limit of $J^{1/2}(\mathbf{A}_{22} - \mathbf{I}_{p-r})$ is a symmetric matrix of independent normal random variables with mean 0 and variance 2 for diagonal and 1 for off-diagonal elements. Because in (iv) $\gamma = lim\ a_n/b_n = lim\ L^{-1/2}$ is $L^{-1/2}$ for case (I) and is 0 for case (II), and because $d_{21}^* = 1$, it follows that $\mathbf{G}_{22}$ in (2.6) is $\mathbf{U}_{22} - L^{-1/2}\mathbf{V}_{22}$ for case (I) and is $\mathbf{U}_{22}$ for case (II). For case (I), the variance is $2L(L-1)$ for diagonal elements of $\mathbf{V}_{22}$ and is $L(L-1)^{-1}$ for off-diagonal elements. Thus, for $\mathbf{G}_{22}$ in case (I), the variance is $2L(L-1)^{-1}$ for diagonal elements and is $L(L-1)^{-1}$ for off-diagonal elements. For case (III), (4.17) holds and $\mathbf{A}^*(n)$ converges to $\mathbf{U}^*$, where given β_j's, $(J-1)\mathbf{U}^*$ is conditionally a noncentral Wishart matrix with covariance matrix block diag$\{0, \mathbf{I}_{p-r}\}$, degrees of freedom $(J-1)$, and noncentrality parameter block diag $\{\sum_{j=1}^{J}(\beta_j-\bar{\beta})(\beta_j-\bar{\beta})',\ \mathbf{0}\}$. Thus, by the standard result on multivariate quadratic forms, the conditional distribution of $(J-1)\mathbf{E}_{22}$ is $W_{p-r}(\mathbf{I}_{p-r},\ J-1-r)$ which is also the unconditional distribution because of nondependence on β_j's.

To obtain the limiting distribution of T in (3.1), observe that for $i = r+1, \ldots, p$,

$\hat{\lambda}_i - 1 = O_P(J^{-1/2})$ for cases (I) and (II) and $\hat{\lambda}_i - 1 = O_P(1)$ for case(III), and that for any positive x

$$\log x = x - 1 - (x-1)^2/2 + O(|x-1|^3).$$

It follows that

$$\begin{aligned}
\frac{(J-1)\hat{\lambda}_i + J(L-1)}{JL-1} - 1 &= \begin{cases} L^{-1}(\hat{\lambda}_i - 1) + O_P(J^{-3/2}) & \text{, for (I),} \\ L^{-1}(\hat{\lambda}_i - 1) + O_P(L^{-2}J^{-3/2}) & \text{, for (II),} \\ J^{-1}(J-1)L^{-1}(\hat{\lambda}_i - 1) + O_P(L^{-2}) & \text{, for (III),} \end{cases} \\
&= \begin{cases} O_P(J^{-1/2}) & \text{, for (I),} \\ O_P(L^{-1}J^{-1/2}) & \text{, for (II),} \\ O_P(L^{-1}) & \text{, for (III).} \end{cases}
\end{aligned}$$

Thus, for $f(\lambda)$ in (3.1),

$$f(\hat{\lambda}_i) = \begin{cases} \frac{1}{2}(L-1)L^{-1}J(\hat{\lambda}_i - 1)^2 + O_P(J^{-1}) & \text{, for (I),} \\ \frac{1}{2}J(\hat{\lambda}_i - 1)^2 + O_P(\max[J^{-1}, L^{-1}]) & \text{, for (II),} \\ (J-1)(\hat{\lambda}_i - 1 - \log\hat{\lambda}_i) + O_P(L^{-1}) & \text{, for (III).} \end{cases}$$

Hence, the limiting distribution of T follows.

To derive the limiting distribution of S for cases (I) and (II), note that

$$\begin{aligned}
d^{-1/2}(S-d) &= (p-r)^{-1/2}(J-1-r)^{-1/2}[(J-1)\sum_{i=r+1}^{p}(\hat{\lambda}_i - 1) + r(p-r)] \\
&\to (p-r)^{-1/2}tr(\mathbf{G}_{22}).
\end{aligned}$$

The results follow from the different distributions of $\mathbf{G}_{22}$ for cases (I) and (II).

5. CONCLUSION

The nondegenerate limiting distribution of the roots of a certain type of determinantal equations was derived under the assumption allowing possibly different convergence rates for the parts of a matrix involved, for the two matrices involved, and for the roots. This general result was applied to the multivariate covariance component problem.

REFERENCES CITED

Amemiya, Y. (1990). A note on the limiting distribution of certain characteristic roots. Statis. Probab. Lett. 9, 465-470.

Amemiya, Y., T. W. Anderson and P.A.W. Lewis (1990). Percentage points for a test of rank in multivariate components of variance. Biometrika 77, 637-641.

Amemiya, Y. and W. A. Fuller (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. Ann. Statist. 12, 497-509.

Anderson, B., T. W. Anderson and I. Olkin (1986). Maximum liklihood estimators and liklihood ratio criteria in multivariate components of variance. Ann. Statist. 14, 405-417.

Anderson, T. W. (1951). The asymptotic distribution of certain characteristic roots and vectors. In: J. Neyman, ed., Proc. 2nd Berkeley Symp. on Math. Statist. and Probab. (Univ. of California Press, Berkeley, CA) pp. 103-130.

Anderson, T. W. (1963). Asymptotic theory for principal components analysis. Ann. Math. Statist. 34, 122-148.

Anderson, T. W. (1984a). Estimating linear statistical relationships. Ann. Statist. 12, 1-45.

Anderson, T. W. (1984b). An Introduction to Multivariate Statistical Analysis. (Wiley, New York, 2nd ed.).

Anderson, T. W. (1989a). The asymptotic distribution of characteristic roots and vectors in multivariate components of variance. In: L.J. Gleser eds., Contributions to Probability and Statistics: Essays in honor of Ingram Olkin (Springer, New York) pp. 177-196.

Anderson, T. W. (1989b). The asymptotic distribution of the liklihood ratio criterion for testing rank in multivariate components of variance. J. Multivariate Anal. 30, 72-79.

Anderson, T. W. and Y. Amemiya (1991). Testing dimensionality in the multivariate analysis of variance. Statist. Probab. Lett. 12, 445-463.

Davis, A. W. (1977). Asymptotic theory for principal component analysis: Nonnormal case. Austral. J. Statist. 19, 206-212.

Eaton, M. l. and D. E. Tyler (1991). On Wiedlant's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. Ann. Statist. 19, 260-271.

Fang, C. and P. R. Krishnaiah (1982). Asymptotic distributions of functions of the eigenvalues of some matrices for nonnormal populations. J. Multivariate Anal. 12, 39-63.

Fujikoshi, Y. (1980). Asynptotic expansion for the distributions of the sample roots under nonnormality. Biometrika 67, 45-51.

Hsu, P. L. (1941). On the limiting distribution of roots of a determinantal equation. J. London Math. Soc. 16, 183-194.

Schott, J. R. and J. G. Saw (1984). A multivariate one-way classification model with random effects. J. Multivariate Anal. 15, 1-12.

Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. Biometrika 63, 639-645.

PAPER II.

TESTS OF RANK FOR COVARIANCE COMPONENTS

ABSTRACT

For the general multivariate mixed effect model, the problem of testing a hypothesis on the rank of a covariance component is discussed. The rank testing problem can be considered to be a multivariate extension of the univariate testing problem for the existence of a random effect. A class of easily computable test statistics is introduced. To develop test procedures useful for various practical situations, asymptotic theory is developed under a wide range of conditions. Test procedures which can be justifiably used for a broad range of applications are derived. For a simple special case, some exact tests are discussed. A simulation study assessing the usefulness of the procedures is presented.

1. INTRODUCTION

The univariate mixed effect analysis has been used widely in biological sciences (animal breeding) and social sciences (panel data, cross-section time-series analysis). For reviews, see e.g., Henderson (1973), Harville (1977), Greene (1990), Robinson (1991), and Searle et al. (1992). When two or more response variables are measured, it is natural to consider that each of the random effects in the model has the same number of variables as the response. See, e.g., Thompson (1973), Amemiya (1985), and Meyer (1985). A general multivariate mixed effects linear model for p-dimensional responses is

$$\mathbf{Y} = \mathbf{X}\mathbf{\Gamma} + \mathbf{Z}\mathbf{B} + \mathbf{E}, \tag{1.1}$$

where $\mathbf{Y}$ is the $N \times p$ matrix of response observations, $N \times k$ $\mathbf{X}$ and $N \times n$ $\mathbf{Z}$ are given explanatory or incidence matrices with the elements of $\mathbf{Z}$ consisting of zero-one classification variables, $\mathbf{\Gamma}$ is a $k \times p$ matrix of unknown parameters, the rows of $n \times p$ random effect matrix $\mathbf{B}$ and $N \times p$ error matrix $\mathbf{E}$ are $1 \times p$ independent random vectors with mean zero. Each row of $\mathbf{E}$ has a $p \times p$ error covariance matrix Σ_{ee}. The random effects are classified into t main effects and interactions. Correspondingly, the n rows of $\mathbf{B}$ can be grouped, where the rows within the $l-th$ group have common covariance matrix Σ_{ll}, $l = 1, 2, \ldots, t$. Thus, model (1.1) contains $t + 1$ covariance components.

A question of interest is whether a particular random effect is significant. The hypothesis of no $l-th$ random effect is that $\Sigma_{ll} = 0$ or equivalently rank $\Sigma_{ll} = 0$. Unlike the univariate case, the existence of an effect can occur in various forms. The random effect may be concentrated in some of the p variables or may be restricted to a space of dimension lower than p. Thus the existence of the $l-th$ random effect consists of p hypotheses, that is rank $\Sigma_{ll} = m$, $m = 1, 2, \ldots, p$. In this sense, testing for the rank of a covariance component is a natural multivariate extension of testing for the existence of random effect in the univariate model. Besides providing information on the structure of random effects, testing for the rank of a covariance component has other consequences in fitting of model (1.1). Properties of estimators of covariance components as well as the fixed effect parameter $\mathbf{\Gamma}$ depend on the ranks of random effect covariance matrices. Knowledge on the rank can be used to develop more efficient estimators than those without using such knowledge. Also, certain inference procedures with incorrectly specified rank or without knowledge of rank may be incorrect. Since tests of rank should be performed in the model building stage, test procedures not involving intensive computing are of practical interest. In this paper, we develop and discuss such test procedures.

First we consider the one-way random effect structure with $t = 1$ in model (1.1), where there is only one unknown covariance matrix other than Σ_{ee}. For such a model, let n be the number of groups corresponding to the one-way classification, and let $\mathbf{y}_{ij}$ be the $p \times 1$ response vector for the $j-th$ individual in the $i-th$ group, $j = 1, 2, \ldots, r_i$, where the number of replicates r_i in general differs over $i = 1, 2, \ldots, n$. Write

$$\mathbf{y}_{ij} = \mathbf{\Gamma}'\mathbf{x}_{ij} + \mathbf{b}_i + \mathbf{e}_{ij}, \; i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, r_i, \tag{1.2}$$

where $\mathbf{b}_i$ and $\mathbf{e}_{ij}$ are $p \times 1$ random vectors with mean zero, and covariance components

$\Sigma_{bb} = \text{Var}\{\mathbf{b}_i\}$ and $\Sigma_{ee} = \text{Var}\{\mathbf{e}_{ij}\}$. Equation (1.2) corresponds to each row of the one-way special case of model (1.1). We consider testing the null hypothesis

$$\text{rank } \mathbf{\Sigma}_{bb} \leq m, \tag{1.3}$$

vs. the alternative that rank $\mathbf{\Sigma}_{bb} > m$, where $0 \leq m < p$. (The two sided problem of rank $\mathbf{\Sigma}_{bb} = m$ vs. rank $\mathbf{\Sigma}_{bb} \neq m$ will be discussed briefly in section 6.) In section 2 a class of test statistics is proposed. For this class, asymptotic null distributions are derived in section 3, assuming that the $\mathbf{e}_{ij}$'s are normally distributed but $\mathbf{b}_i$'s can have any distribution. Some care is required in considering asymptotics in covariance component type problems. See Miller (1977). For model (1.2), we need to consider different situations depending on whether n and/or r_i's are considered large. In a typical animal breeding study, n is large but r_i's are small. For economic and social science problems, various situations can arise, but a case with small n and large r_i's is possible. Such consideration is given in section 3, and asymptotics are discussed under three different sets of assumptions concerning n and r_i's. Test procedures justifiable for all three cases are derived. As will be seen in section 5, such asymptotic theory has direct application in the more general model of type (1.1) than the one-way model (1.2). In section 4, exact finite sample tests are discussed for a special case of model (1.2) with no $\mathbf{x}_{ij}$, common $r_i = r$, and the normality of both $\mathbf{b}_i$'s and $\mathbf{e}_{ij}$'s. Extensions to models more complicated than the one-way case (1.2) are considered in section 5. Section 6 reports on a simulation study comparing different tests given in section 2. Derivations are given in section 7.

Tests of the rank of the between-group covariance matrix in the one-way classification model have been discussed in Amemiya and Fuller (1984), Schott and Saw (1984), Anderson et al. (1986), Anderson (1989), Amemiya et al. (1990), and Anderson and Amemiya (1991). The relevance of these references is mentioned in various parts of the following sections.

2. TEST STATISTICS

As mentioned in the previous section, we consider test statistics which can be computed readily without involved computations. We use the fitting-constant technique (the fixed effect MANOVA technique). The one-way model (1.2) can be expressed in the matrix form (1.1) with $\mathbf{Z} =$ block diag$\{\mathbf{1}_{r_1}, \mathbf{1}_{r_2}, \ldots, \mathbf{1}_{r_n}\}$, where $\mathbf{1}_a$ denotes the $a \times 1$ vector of one's. For any matrix $\mathbf{A}$, we denote the projection matrices

$$\begin{aligned} \mathbf{P_A} &= \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}', \\ \mathbf{M_A} &= \mathbf{I} - \mathbf{P_A}. \end{aligned} \tag{2.1}$$

Then, the $p \times p$ fitting-constant mean square matrix for the group-effect is

$$\begin{aligned} \mathbf{m}_{bb} &= \frac{1}{d_b}\mathbf{Y}'(\mathbf{P}_{\mathbf{X},\mathbf{Z}} - \mathbf{P_X})\mathbf{Y} \\ &= \frac{1}{d_b}\mathbf{Y}'\mathbf{P}_{\mathbf{M_X Z}}\mathbf{Y}, \end{aligned} \tag{2.2}$$

and the $p \times p$ error mean square matrix is

$$\mathbf{m}_{ee} = \frac{1}{d_e}\mathbf{Y}'\mathbf{M}_{\mathbf{X},\mathbf{Z}}\mathbf{Y}, \tag{2.3}$$

where $d_b =$ rank $(\mathbf{X}, \mathbf{Z})-$ rank $\mathbf{X}$, $d_e = N-$ rank $(\mathbf{X}, \mathbf{Z})$, and $N = \sum_{i=1}^{n} r_i$. Note that

$$\begin{aligned} E\{\mathbf{m}_{bb}\} &= q\boldsymbol{\Sigma}_{bb} + \boldsymbol{\Sigma}_{ee}, \\ E\{\mathbf{m}_{ee}\} &= \boldsymbol{\Sigma}_{ee}, \end{aligned} \tag{2.4}$$

where $q = \frac{1}{d_b}\{\sum_{i=1}^{n} r_i - tr[(\mathbf{X}'\mathbf{X})^{-1}\sum_{i=1}^{n} r_i^2 \bar{\mathbf{x}}_{i.}\bar{\mathbf{x}}_{i.}']\}$, and $\bar{\mathbf{x}}_{i.} = \frac{1}{r_i}\sum_{j=1}^{r_i}\mathbf{x}_{ij}$.

The condition that rank $\Sigma_{bb} = m$ is equivalent to the condition that exactly $p - m$ of the p roots of $\left|(q\Sigma_{bb} + \Sigma_{ee}) - \lambda\Sigma_{ee}\right| = 0$ are equal to one and the m largest roots are larger than one. Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$ be the ordered roots of the sample determinantal equation

$$\left|\mathbf{m}_{bb} - \hat{\lambda}\mathbf{m}_{ee}\right| = 0. \tag{2.5}$$

The above discussion based on the expected determinantal equation suggests that we should reject the null (1.3) if $\hat{\lambda}_i$'s, $i = m+1, m+2, \ldots, p$, are large in some sense. We consider a class of statistics of the form

$$T = \sum_{i=m+1}^{m+a} g(\hat{\lambda}_i, \frac{d_b}{d_e}), \tag{2.6}$$

where an integer $a = 1, 2, \ldots, p - m$ decides the number of $\hat{\lambda}_i$'s to be included. We assume that the integer a and the form of the funcion $g(.,.)$ are free of the data, n, and r_i's. Note that d_b and d_e enter the statistic T only through the ratio d_b/d_e. This assumption is convenient in developing a unified theory and is satisfied by all practical forms of the statistic. For the function $g(\lambda,\ \delta)$, we consider two types, both nondecreasing in λ. The first type simply uses a smooth function increasing at 1;

Type 1: $g(\lambda,\ \delta)$ is defined on $\mathbf{\Omega} = \{(\lambda, \delta), \lambda > 0, \delta \geq 0\}$ such that $\frac{\partial g}{\partial \lambda}(\lambda, \delta)$ is continuous and nonnegative on $\mathbf{\Omega}$, and that for every $\delta \geq 0$, $\frac{\partial g}{\partial \lambda}(1, \delta) > 0$.

For the second type, we measure the distance between λ and 1 using some distance function, and truncate it to be zero for $\lambda \leq 1$, considering that $\hat{\lambda}_i \leq 1,\ i = m + 1, m+2, \ldots, p$, may not provide information on (1.3);

Type 2: $g(\lambda, \delta) = f(\lambda, \delta)I(\lambda > 1)$, where $I(.)$ is the indicator function, and the distance function $f(\lambda, \delta)$ defined on Ω satisfies that for every $\delta \geq 0$

$$\begin{aligned} f(1,\delta) &= 0, \\ \frac{\partial f}{\partial \lambda}(\lambda,\delta) &< 0, \quad 0 < \lambda < 1, \\ &= 0, \quad \lambda = 1, \\ &> 0, \quad \lambda > 1, \\ \frac{\partial^2 f}{\partial \lambda^2}(1,\delta) &> 0, \end{aligned}$$

and that $\frac{\partial^2 f}{\partial \lambda^2}(\lambda, \delta)$ is continous on Ω.

Some particular forms of the statistic T are obtained by choosing a particular a and a particular $g(.,.)$. Some choices of a and $g(.,.)$, can be suggested either by considering analogies of those used in a similar problem or by extending one for a simple special case of our model. In the (fixed effect) multivariate regression, a general linear hypothesis can be tested using four statistics which are functions of the roots of a determinantal equation similar to (2.5). See, e.g., Anderson (1984). By analogy, we consider, for our problem (1.3),

$$\begin{aligned} T_W &= \sum_{i=m+1}^{p} \frac{d_e}{d_b} log(1 + \frac{d_b}{d_e}\hat{\lambda}_i), \\ T_H &= \sum_{i=m+1}^{p} \hat{\lambda}_i, \\ T_B &= \sum_{i=m+1}^{p} \hat{\lambda}_i(1 + \frac{d_b}{d_e}\hat{\lambda}_i)^{-1}, \\ T_R &= \hat{\lambda}_{m+1}, \end{aligned} \tag{2.7}$$

where these are adaptations of Wilks criterion, Hotelling-Lawley trace, Bartlett trace, and Roy largest root, respectively. Note that these four statistics have $g(.,.)$ of Type 1 with $a = p - m$ for T_W, T_H, and T_B, and $a = 1$ for T_R. Note that $g(\lambda, 0)$ for

T_W is defined as $\lim_{\delta \to 0^+} g(\lambda, \delta) = \lambda$ to satisfy Type 1 condition. Schott and Saw (1984) showed that T_R is the likelihood ratio test statistic for testing the null (1.3) vs. the alternative that rank $\Sigma_{bb} = m + 1$ for the balanced normal random model (1.2) with $k = 1, \mathbf{x}_{ij} = 1$ for all i and j, normally distributed $\mathbf{b}_i$ and $\mathbf{e}_{ij}$, and $r_i = r$ for all i. For this balanced normal model, Anderson et al. (1986) derived the likelihood ratio test statistic for testing problem (1.3). Anderson (1989) derived the limiting null distribution for fixed r as $n \to \infty$, and Amemiya et al. (1990) gave a table of percentiles of such (non-χ-squared) limiting distribution. Anderson and Amemiya (1991) suggest the use of an extension of this statistic for more general models. The extension is

$$T_L = \sum_{i=m+1}^{p} f(\hat{\lambda}_i, \frac{d_b}{d_e}) I(\hat{\lambda}_i > 1), \tag{2.8}$$

where $I(.)$ is the indicator function, and

$$f(\lambda, \delta) = -log\lambda + (1 + \frac{1}{\delta}) log(\frac{\delta\lambda + 1}{\delta + 1}).$$

Note that T_L has $a = p - m$ and $g(\lambda, \delta)$ of Type 2 with $f(\lambda, 0) = lim_{\delta \to 0^+} f(\lambda, \delta)$ $= -log\lambda + \lambda - 1$. For the general model (1.2), T_L is not the likelihood ratio test statistic (even for normal $\mathbf{b}_i$ and $\mathbf{e}_{ij}$), but can easily be computed for different values of m without iterative model fitting. Another possible statistic using Type 2 function is based on the squared Euclidian distance

$$T_S = \sum_{i=m+1}^{m+a} \frac{1}{2}(\hat{\lambda}_i - 1)^2 I(\hat{\lambda}_i > 1). \tag{2.9}$$

These particular statistics are used as examples later, and some comparison will be discussed. But, for developing theoretical results, we consider the general statistic T given by (2.6) with either Type 1 or 2 $g(\lambda, \delta)$ function. Such consideration provides a convenient unified approach to developing asymptotic and exact distribution theory for various possible statistics. Finding an optimal choice in some sense among statistics of the form T is not our focus here.

Note that the testing problem (1.3) and the roots $\hat{\lambda}_i$'s of (2.5) are invariant under transformation $\mathbf{YC}'$, i.e., $(\mathbf{C}\mathbf{\Sigma}_{bb}\mathbf{C}', \mathbf{C}\mathbf{\Sigma}_{ee}\mathbf{C}')$, with $p \times p$ nonsingular $\mathbf{C}$. Thus, to discuss distributional properties of the test statistic T, we can assume, without loss of generality, that $\mathbf{\Sigma}_{bb}$ is diagonal and $\mathbf{\Sigma}_{ee} = \mathbf{I}_p$.

3. ASYMPTOTIC NULL DISTRIBUTIONS

To develop practical asymptotic theory for model (1.2), we consider different situations depending on the number of groups n and the replicate numbers r_i. Case I corresponds to large n and small r_i's. For case II both n and r_i's are large, and case III represents the situation with small n and large r_i's. We use r to denote an imaginary index representing "largeness" of r_i's, and our three cases are formally defined as

case I. $n \to \infty$ and r fixed,

case II. $n \to \infty$ and $r \to \infty$,

case III. n fixed and $r \to \infty$.

Technical meaning of these three cases are explicitly given in the following assumptions. For all three cases we assume

(a) The $\mathbf{b}_i$'s are independent and identically distributed with mean zero, covariance matrix block diag $\{\mathbf{D}, \mathbf{0}\} = \text{diag}\{\delta_1, \ldots, \delta_{m_0}, 0, \ldots, 0\}$, where $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_{m_0} > 0$, and finite fourth moments, independently from $\mathbf{e}_{ij}$'s which are independent $N_p(\mathbf{0}, \mathbf{I}_p)$ random vectors,

(b) The number k of columns of $\mathbf{X}$ is fixed, and d_b defined in (2.2) can be written as $n - l$ for an l fixed over the limiting sequence,

(c) Over the limiting sequence,

$$\begin{aligned}
\frac{1}{nr}\sum_{i=1}^{n} r_i &\to c_1, \\
\frac{1}{nr^2}\sum_{i=1}^{n} r_i^2 &\to c_2, \\
\left(\frac{1}{nr}\mathbf{X}'\mathbf{X}\right)^{-1} &= O(1), \\
\frac{1}{nr^2}\sum_{i=1}^{n} r_i^2 \bar{\mathbf{x}}_{i.}\bar{\mathbf{x}}_{i.}' &= O(1),
\end{aligned}$$

where $0 < c_s < \infty$, $s = 1,\ 2$.

Note that in assumption (a) we assumed rank $\Sigma_{bb} = m_0$ and already incorporated the canonical transformation based on invariance to have diagonal Σ_{bb} and Σ_{ee}. The distribution of $\mathbf{b}_i$'s is unspecified in (a) except for the existence of fourth moments. Assumption (b) is not a real constraint, and simply assumes that the column space of $\mathbf{X}$ does not change over the sequence. For a practical model with observational explanatory variables, $l = 0$ without intercept and $l = 1$ with intercept in the model. Assumption (c) explains what is actually assumed for n, r and r_i's in cases I, II, and III. Under (c), $\sum_{i=1}^{n} r_i^s = O(nr^s)$ for $s = 1, 2$, where r can be taken to be one for case I and n is a fixed constant for case III. Intuitively, for cases II and III, most r_i's are of order r but some r_i's can be small. The last two conditions in (c) are on the behavior of $\mathbf{x}_{ij}$'s in the sequence. These conditions are satisfied, for example, if components of $\mathbf{x}_{ij}$'s are variables such as intercept and observational covariates behaving similar to independent and identically distributed random variables. Classification dummy variables as part of $\mathbf{x}_{ij}$'s may or may not satisfy assumptions (b) and (c) depending on the situation. For example, consider the balanced two-way mixed effect model, where each combination of u levels of a fixed effect and n levels of a random effect is replicated v times. Then, model (1.2) holds with $r_i = uv$ for $i = 1, 2, \ldots, n, \mathbf{X} = \mathbf{1}_n \otimes \mathbf{I}_u \otimes \mathbf{1}_v$, and $k = u$. Since (b) assumes fixed k, our asymptotics do not include the case with large u, the number of levels for the fixed effect. For fixed u, noting that $\sum_{i=1}^{n} r_i^2 \bar{\mathbf{x}}_{i.} \bar{\mathbf{x}}_{i.}' = nv^2 \mathbf{1}_n \mathbf{1}_n'$ and $\mathbf{X}'\mathbf{X} = nv\mathbf{I}_u$, all four conditions in (c) are satisfied with $r = v$. Depending on whether n and/or v are large, we can have cases I, II, or III.

Before we discuss asymptotics for the test statistics, we begin by deriving the limiting distributions of $\hat{\lambda}_{m+1} \geq \cdots \geq \hat{\lambda}_p$, the $p - m$ smaller roots of (2.5). Note that the null (1.3) is a composite hypothesis and that the true rank m_0 of $\mathbf{\Sigma}_{bb}$ may be $0, 1, \ldots, m$ under the null. The limiting distribution of $\hat{\lambda}_i$'s, $i = m + 1, \ldots, p$, depends on the true rank $m_0 \leq m$. To develop a cut-off point for the statistic T which gives a specified asymptotic size, we need to consider $m + 1$ different limiting distributions of $\hat{\lambda}_i$'s depending on $m_0 = 0, 1, 2, \ldots, m$. For a given m_0, we first derive the limiting distribution of $\hat{\lambda}_i$'s, $i = m_0 + 1, \ldots, p$ (a possibly larger set than $\hat{\lambda}_i$'s, $i = m + 1, \ldots, p$). A difficulty associated with the derivation is that parts of $\mathbf{m}_{bb}$ have different rates of convergence for cases II and III. To overcome this difficulty, a nonstandrd result on the limiting distribution of roots will be utilized. See section 7 for details. For cases I and II, we have the following result. All derivations are given in section 7.

Theorem 1. For cases I and II, under model (1.2) with assumptions (a), (b), and (c), $\sqrt{n}(\hat{\lambda}_{m_0+1} - 1, \hat{\lambda}_{m_0+2} - 1, \ldots, \hat{\lambda}_p - 1)$ converges in distribution to the ordered roots $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{p-m_0}$ of a $(p - m_0) \times (p - m_0)$ symmetric matrix $\mathbf{G}(m_0)$, where the functionally independent elements of $\mathbf{G}(m_0)$ are independent normal random variables with mean zero and variance $2\sigma^2$ for diagonal and σ^2 for off-diagonal elements, $\sigma^2 = c_1(c_1 - 1)^{-1}$ for case I and 1 for case II, and where it is understood that for case I, c_1 is defined in (c) with $r = 1$.

For case III, the limiting distribution of $\hat{\lambda}_i$'s $i = m + 1, \ldots, p$ takes a slightly different form and requires an additional assumption. Namely, we assume

(d) $n - l \geq p$ and $\frac{1}{r}\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{Z} \rightarrow \mathbf{K}$ for some $\mathbf{K}$ of rank $\geq p$.

For case III, n is fixed, and needs to be assumed large enough to obtain a nondegenerate distribution of $\hat{\lambda}_i$'s. The second condition in (d) is

$$\frac{1}{r}[\mathbf{D}(r_i) - \mathbf{D}(r_i)\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'\mathbf{D}(r_i)] \rightarrow \mathbf{K}, \tag{3.1}$$

where $\mathbf{D}(r_i) = \operatorname{diag}\{r_1, r_2, \ldots, r_n\}$, and $\mathbf{X}_0 = (\bar{\mathbf{x}}_{1.}, \bar{\mathbf{x}}_{2.}, \ldots, \bar{\mathbf{x}}_{n.})$. Under assumption (c), (3.1) is nearly equivalent to assuming that $r^{-1}r_i \rightarrow r_i^0, i = 1, 2, \ldots, n$, but (3.1) is the exact condition required for the following Theorem 2. The rank condition on $\mathbf{K}$ is rather weak, because assumption (b) assumes that for every r rank $\frac{1}{r}\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{Z} = n - l \geq p$.

Theorem 2. For case III, under model (2) with the assumptions (a), (b), (c), and (d), $\hat{\lambda}_i$, $i = m_0 + 1, m_0 + 2, \ldots, p$ converge in distribution to the ordered roots $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_{p-m_0}$ of a $(p - m_0) \times (p - m_0)$ matrix $\mathbf{G}(m_0)$, where $d_b\mathbf{G}(m_0)$ has the Wishart distribution with covariance matrix $\mathbf{I}_{p-m_0}$ and degrees of freedom $n - l - m_0$.

Now, we derive the limiting distributions of the general statistic T under each value of the true rank $m_0 = 0, 1, \ldots, m$. For Type 1 statistics, we have the following result.

Theorem 3. Let model (1.2) hold with assumptions (a), (b), and (c), and let T in (2.6) have $g(\lambda, \delta)$ of Type 1. Suppose that rank $\Sigma_{bb} = m_0 \leq m$.

- For case I,
$$\sqrt{n}[T - ag(1, \frac{d_b}{d_e})] \xrightarrow{L} M_1(\mathrm{I}, m_0),$$
where
$$M_1(\mathrm{I}, m_0) = \frac{\partial g}{\partial \lambda}(1, (c_1 - 1)^{-1}) \sum_{i=m-m_0+1}^{m-m_0+a} \gamma_i,$$
c_1 is given in (c) with $r = 1$, and γ_i's are given in Theorem 1 for case I.

- For case II,
$$\sqrt{n}[T - ag(1, \frac{d_b}{d_e})] \xrightarrow{L} M_1(\mathrm{II}, m_0),$$
where
$$M_1(\mathrm{II}, m_0) = \frac{\partial g}{\partial \lambda}(1, 0) \sum_{i=m-m_0+1}^{m-m_0+a} \gamma_i,$$
and γ_i's are given in Theorem 1 for case II.

- For case III with assunption (d),
$$T \xrightarrow{L} M_1(\mathrm{III}, m_0),$$
where
$$M_1(\mathrm{III}, m_0) = \sum_{i=m-m_0+1}^{m-m_0+a} g(\gamma_i, 0),$$
and γ_i's are given in Theorem 2.

For the statistic T with $g(\lambda, \delta)$ of Type 2, we have the following limiting distribution results for each $m_0 \leq m$.

Theorem 4. Let model (1.2) hold with the assumptions (a), (b), and (c), and let T in (2.6) have $g(\lambda, \delta)$ of Type 2. Suppose that rank $\Sigma_{bb} = m_0 \leq m$.

- For case I,

$$nT \xrightarrow{L} M_2(\mathrm{I}, m_0),$$

where

$$M_2(\mathrm{I}, m_0) = \frac{\partial^2 f}{\partial \lambda^2}(1, (c_1 - 1)^{-1}) \sum_{1=m-m_0+1}^{m-m_0+a} \gamma_i^2 I(\gamma_i > 0),$$

c_1 is given in (c) with $r = 1$, and γ_i's are given in Theorem 1 for case I.

- For case II

$$nT \xrightarrow{L} M_2(\mathrm{II}, m_0),$$

where

$$M_2(\mathrm{II}, m_0) = \frac{\partial^2 f}{\partial \lambda^2}(1, 0) \sum_{i=m-m_0+1}^{m-m_0+a} \gamma_i^2 I(\gamma_i > 0),$$

and γ_i's are given in Theorem 1 for case II.

- For case III with assumption (d),

$$T \xrightarrow{L} M_2(\mathrm{III}, m_0),$$

where

$$M_2(\mathrm{III}, m_0) = \sum_{i=m-m_0+1}^{m-m_0+a} f(\gamma_i, 0) I(\gamma_i > 1),$$

and γ_i's are given in Theorem 2.

Theorems 3 and 4 have shown that, for each of Types 1 and 2, and each of cases I, II, and III, T with normalizing constants free of m_0 has a limiting distribution depending on $m_0 \leq m$. Thus, to obtain cut-off points making the size asymptotically equal to a given level α, we must find k_α such that, for given Type $l = 1, 2$ and case h=I, II, III,

$$\sup_{0 \le m_0 \le m} P\{M_l(h, m_0) > k_\alpha\} = \alpha.$$

The next theorem characterizes such a k_α, using the result that $M_l(h, m)$ is stochastically larger than $M_l(h, m_0)$ for $m_0 = 0, 1, \ldots, m-1$.

Theorem 5. For $M_l(h, m_0)$ as given in Theorems 3 and 4

$$\sup_{0 \le m_0 \le m} P\{M_l(h, m_0) > k_\alpha(l, h)\} = \alpha.$$

where $k_\alpha(l, h)$ is the upper α point of $K(l, h)$ such that

$$
\begin{aligned}
K(1,\mathrm{I}) &= \frac{\partial g}{\partial \lambda}(1, \frac{1}{c_1 - 1}) \frac{c_1^{1/2}}{(c_1 - 1)^{1/2}} H_1, \\
K(1,\mathrm{II}) &= \frac{\partial g}{\partial \lambda}(1, 0) H_1, \\
K(1,\mathrm{III}) &= \sum_{i=1}^{a} g(\gamma_i, 0) \\
K(2,\mathrm{I}) &= \frac{\partial^2 f}{\partial \lambda^2}(1, (c_1 - 1)^{-1}) \frac{c_1}{(c_1 - 1)} H_2, \\
K(2,\mathrm{II}) &= \frac{\partial^2 f}{\partial \lambda^2}(1, 0) H_2, \\
K(2,\mathrm{III}) &= \sum_{i=1}^{a} f(\gamma_i, 0) I(\gamma_i > 1),
\end{aligned}
$$

where c_1 is as given in (c) with $r = 1$, γ_i's are as given in Theorem 2 for case III with $m_0 = m$,

$$
\begin{aligned}
H_1 &= \sum_{i=1}^{a} \eta_i, \qquad (3.2) \\
H_2 &= \sum_{i=1}^{a} \eta_i^2 I(\eta_i > 0),
\end{aligned}
$$

and the η_i's are the eigenvalues of a $(p-m) \times (p-m)$ symmetric matrix $\mathbf{F}_0$ with normal independent $(p-m) \times (p-m+1)/2$ elements with mean 0 and variance 2 for diagonal and 1 for off-diagonal.

Since $c_1 = lim\ n^{-1} \sum_{i=1}^{n} r_i$ for case I (with $r = 1$), Theorems 3, 4, and 5 characterize the required asymptotic cut-off points for each given T and for each of cases I, II, and III. But, in practice, it may not be very clear which of cases I, II, or III a given data configuration corresponds to. Thus, it seems important to develop a test procedure which is appropriate under a wide range of situations. Ideally we wish to develop a test procedure which can be justified under all cases I, II, and III and which uses readily available cut-off points. Such a test could be used in practice without worrying about the sizes of n and r_i's. As seen in Theorem 5, the distributions used for the asymptotic cut-off points are similar for cases I and II, and hence, with properly adjusting normalizing constants, we can easily develop test procedures justifiable under cases I and II.

Theorem 6. Let model (1.2) hold with assumptions (a), (b), and (c). For $l = 1, 2$, let $H_{l\alpha}$ denote the upper α point of H_l in (3.2). For T in (2.6) with Type 1 $g(\lambda,\ \delta)$, let

$$T_1 = \sqrt{\frac{d_b d_e}{d_b + d_e}} \left\{ \frac{\partial g}{\partial \lambda}(1,\ \frac{d_b}{d_e}) \right\}^{-1} [T - ag(1,\ \frac{d_b}{d_e})].$$

For T in (2.6) with Type 2 $g(\lambda,\ \delta)$, let

$$T_2 = \frac{d_b d_e}{d_b + d_e} \left\{ \frac{\partial^2 f}{\partial \lambda^2}(1,\ \frac{d_b}{d_e}) \right\}^{-1} T.$$

Then, for each of cases I and II, and for $l = 1,\ 2$,

$$\sup_{H_0} P\{T_l > H_{l\alpha}\} \to \alpha,$$

where H_0 is given in (1.3).

Thus, the test procedures based on T_1 and T_2 with any $g(\lambda,\ \delta)$ function (satisfying Type 1 or 2 condition) can be used in practice, if n is considered large, whether or not the r_i's are large, provided that the cut-off points $H_{l\alpha}$ can be obtained. For a particular value of $a = 1, 2, \ldots, p-m$, the distribution of H_l is simple, and is tabulated. For $a = p - m$, i.e., for T using the $(p-m)$ smaller roots $\hat{\lambda}_i$'s , $H_1 \sim N(0, 2(p-m))$, and the percentage points of H_2 are tabulated in Amemiya et al. (1990). For $a = 1$, the distribution of largest root of the symmetric normal matrix can be used to obtain percentiles of H_1 and H_2.

Developing a test procedure which can be used under case III as well as cases I and II is less straightforward, because of the dependency of $K(l,\ \mathrm{III})$ on a particular function $g(\lambda,\ \delta)$ and n. However, we can characterize a procedure justifiable under all cases I, II, and III. Our idea for developing such a test was to consider the limiting behavior of a case III cut-off point (not a statistic) under conditions of cases I and II, and to find appropriate normalizing constants.

Theorem 7. Let model (1.2) hold with assumptions (a), (b), (c), and (d). For T in (2.6) with Type 1 $g(\lambda,\ \delta)$, let

$$T_1^* = \sqrt{\frac{d_e}{d_b + d_e}\frac{\partial g}{\partial \lambda}(1,0)\left\{\frac{\partial g}{\partial \lambda}(1, \frac{d_b}{d_e})\right\}^{-1}}\ [T - ag(1, \frac{d_b}{d_e})] + ag(1,0),$$

and let $k_{1\alpha}$ be the upper α point of $K(1,\ \mathrm{III})$ specified in Theorem 5. For T in (2.6) with Type 2 $g(\lambda,\ \delta)$, let

$$T_2^* = \frac{d_e}{d_b + d_e}\frac{\partial^2 f}{\partial \lambda^2}(1,\ 0)\left\{\frac{\partial^2 f}{\partial \lambda^2}(1,\ \frac{d_b}{d_e})\right\}^{-1} T$$

and let $k_{2\alpha}$ be the upper α point of $K(2,\ \mathrm{III})$ specified in Theorem 5. Then, for each of cases I, II, and III, for $l = 1, 2$, and for $0 < \alpha < 1$,

$$\sup_{H_0} P\{T_l^* > k_{l\alpha}\} \to \alpha,$$

where H_0 is given in (1.3).

Hence, for any function $g(\lambda,\ \delta)$ of Types 1 and 2, we can use the modified statistic T_l^* and the cut-off points $k_{l\alpha}$ to test the null (1.3) in practice without worrying whether n is large r_i is large or both are large. We note that for $l = 1, 2$, $k_{l\alpha}$ is the upper α point of the distribution of $K(l,\ \mathrm{III})$ corresponding to case III in Theorem 5. Such a distribution is rather complex in general because of the dependency on $g(\lambda,\ \delta)$ and the γ_i's, roots of a Wishart matrix. However, for some choice of $g(\lambda,\ \delta)$ and a, the cut-off points $k_{l\alpha}$ are available. We modify T_H in (2.7) as

$$T_H^* = \sqrt{\frac{d_e}{d_b + d_e}} \sum_{i=m+1}^{p} (\hat{\lambda}_i - 1) + (p - m). \tag{3.3}$$

For this statistic, $g(\lambda,\ \delta) = \lambda$ free of δ simplifies the expression of T_H^*, and gives

$$\begin{aligned} k_{1\alpha} &= \mathrm{tr}\left\{\frac{1}{d_b} W_{p-m}((\mathbf{I}_{p-m},\ n - l - m)\right\}_\alpha \\ &= \frac{1}{d_b}\chi^2_{(p-m)(n-l-m),\ \alpha}. \end{aligned}$$

Thus, T_H^* and a table of chi-square cut-off points can be used to test the null (1.3) without much restrictions on the configurations of n and r_i's and on the distribution of the random effect $\mathbf{b}_i$. The χ^2 test for all cases I, II, and III can also be developed easily by modifying T_W and T_B in (2.7) using Theorem 7, because $g(\lambda, 0)$ used in Theorem 5 is λ for T_W and T_B as well. Modifying T_R in (2.7), a statistic

$$T_R^* = \sqrt{\frac{d_e}{d_b + d_e}(\hat{\lambda}_{m+1} - 1) + 1}$$

and the percentiles of the largest root of $W_{p-m}(\mathbf{I}_{p-m}, n - l - m)$ would produce a widely applicable test procedure. For the test statistic T_L in (2.8), the modification is unnecessary, because of the structure of $\frac{\partial^2 f}{\partial \lambda^2}(\lambda, \delta)$, i.e., $T_L^* = T_L$, and $k_{2\alpha}$ is the upper α point of $\sum_{i=1}^{p-m} f(\gamma_i, 0) I(\gamma_i > 1)$ using the roots of $W_{p-m}(I_{p-m}, n - l - m)$. For T_S in (2.9),

$$T_S^* = \frac{d_e}{d_b + d_e}\frac{1}{2}\sum_{i=m+1}^{p}(\hat{\lambda}_i - 1)^2 I(\hat{\lambda}_i > 1),$$

and the upper α point of $\frac{1}{2}\sum_{i=1}^{p-m}(\gamma_i - 1)^2 I(\gamma_i > 1)$ provide an approximate test procedure. But, the cut-off points for T_L^* and T_S^* have not been tabulated.

4. EXACT TESTS

For a special case of model (1.2), we can characterize exact cut-off points for test statistics T in (2.6). Consider the balanced normal model

$$\mathbf{y}_{ij} = \mu + \mathbf{b}_i + \mathbf{e}_{ij},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, r, \tag{4.1}$$

where $\mathbf{b}_i$, in addition to $\mathbf{e}_{ij}$, is assumed to be normally distributed. Note that the only fixed effect is the overall mean and the number r of replicates is common for all groups. For this model, in addition to $\mathbf{m}_{ee}$, $\mathbf{m}_{bb}$ is also a Wishart matrix, i.e.,

$$\begin{aligned} (n-1)\mathbf{m}_{bb} &\sim \mathbf{W}_p(r\mathbf{\Sigma}_{bb} + \mathbf{\Sigma}_{ee},\ n-1), \\ n(r-1)\mathbf{m}_{ee} &\sim \mathbf{W}_p(\mathbf{\Sigma}_{ee},\ n(r-1)). \end{aligned} \tag{4.2}$$

For this situation, using the monotonicity result of Anderson and Das Gupta (1964), Schott and Saw (1984) derived the following result.

Result. Let (4.2) hold with $m_0 = \text{rank } \mathbf{\Sigma}_{bb} \le m$. Let $h(\hat\lambda_{m+1}, \ldots, \hat\lambda_p)$ be any function of $\hat\lambda_i$, $i = m+1, \ldots, p$ of $\left|\mathbf{m}_{bb} - \hat\lambda \mathbf{m}_{ee}\right| = 0$ such that $h(\hat\lambda^*_{m+1}, \ldots, \hat\lambda^*_p) \le h(\hat\lambda_{m+1}, \ldots, \hat\lambda_p)$ for any $\hat\lambda^*_i \le \hat\lambda_i$, $i = m+1, \ldots, p$. Then, for any constant k,

$$\sup_{H_0} P\{h(\hat\lambda_{m+1}, \ldots, \hat\lambda_p) > k\} = P\{h(\nu_{m+1}, \ldots, \nu_p) > k\},$$

where H_0 is given in (1.3), and $\phi_i = \frac{n-1}{n(r-1)}\nu_i$, $i = 1, 2, \ldots, p-m$ are the ordered roots of

$$|\mathbf{W}_1 - \phi \mathbf{W}_2| = 0,$$

$\mathbf{W}_1 \sim \mathbf{W}_{p-m}(\mathbf{I}_{p-m},\ n-1-m)$, $\mathbf{W}_2 \sim \mathbf{W}_{p-m}(\mathbf{I}_{p-m},\ n(r-1))$, and $\mathbf{W}_1$ and $\mathbf{W}_2$ are independent.

Note that our T in (2.6) as a function of $\hat{\lambda}_{m+1}, \ldots, \hat{\lambda}_p$ satisfies the condition on h in the above result. Thus, for this simple special model, we can obtain cut-off points for T using the upper percentiles of the distribution of T with $\hat{\lambda}_i$ replaced by ν_i, $i = m+1, \ldots, p$. For certain choices of a and $g(\lambda, \delta)$, such percentiles are available. We see that the distribution of ν_i's is the same as those appearing in the general linear hypothsis in the multivariate regression. Thus, the four statistics T_W, T_B, T_H, and T_R in (2.7) suggested by analogy to the multivariate regression problem in fact have the same null distributions for the simple random effect model as those for the regression. In using such tabulated percentage points, e.g., as given in Anderson (1984, pp. 609-637), we note that the degrees of freedom are $n-1-m$ and $n(r-1)$. For the multivariate regression, df_1 = the number of coeffiocients being tested and df_2 = the error degrees of freedom. For the regression problems, df_1 is typically small. As a result, the available tables for the cut-off points typically cover only small values of df_1. For the random effect model (4.1), the number of groups n is often large in practice. For such a case, df_1 is large, and existing tables may not contain required cut-off points.

5. EXTENSIONS

In this section we are concerned with developing test procedures for testing the rank of a random effect in multivariate mixed models with more than two covariance components, i.e., $t \geq 2$ in model (1.1). For the one-way madel (1.2), we considered the class of test statistics T in (2.6) based on the roots of the determinatal equation $\left|\mathbf{m}_{bb} - \hat{\lambda}\mathbf{m}_{ee}\right| = 0$ in (2.5), involving two matrices $\mathbf{m}_{bb}$ and $\mathbf{m}_{ee}$. These matrices are the mean square matrices obtained by a partitionning of the total sum of squares using the fitting-constants method. For the general model (1.1), the fitting-constants method can also produce a partition of the total sum of squares with a number of mean square matrices. In applying this method to this general model (1.1), we act as if all effects (other than the errors) are fixed, and fit the covariate $\mathbf{X}$ before the constant $\mathbf{Z}$ corresponding to the random effects. For the general model (1.1), suppose that we are interested in testing for the rank of a particular covariance component $\mathbf{\Sigma}_{ll}$, i.e.,

$$\text{rank } \mathbf{\Sigma}_{ll} \leq m, \qquad 0 \leq m < p.$$

The ordering of the columns of $\mathbf{Z}$ in this fitting-constants technique is not in general unique, although some rules must be followed. For example, an interaction effect must be fitted after the corresponding main effects and lower order interaction effects, and any effect nested in another effect A, say, must be fitted before A. Often, after possibly changing the order of fitting-constants, we can obtain two mean square matrices, say, $\mathbf{m}_{ll}$ and $\mathbf{m}_{ss}$, satisfying the pairwise expected difference restriction

$$\begin{aligned} E\{\mathbf{m}_{ll}\} &= q_l \boldsymbol{\Sigma}_{ll} + \boldsymbol{\Sigma}_{00}, \\ E\{\mathbf{m}_{ss}\} &= \boldsymbol{\Sigma}_{00}, \end{aligned} \tag{5.1}$$

where q_l is a known constant and $\boldsymbol{\Sigma}_{00}$ is a positive definite matrix, usually a linear combination of the covariance components other than $\boldsymbol{\Sigma}_{ll}$. Note that (5.1) has the same form as the expectations of $\mathbf{m}_{bb}$ and $\mathbf{m}_{ee}$ of the one-way model (1.2). Hence, we suggest the use of a test statistic which is a function of the $(p-m)$ smaller roots of the determinantal equation $\left|\mathbf{m}_{ll} - \hat{\lambda}\mathbf{m}_{ss}\right| = 0$. For example, any test statistic T discussed in section 3 can be used, where the degrees of freedom d_b and d_e are replaced by d_l and d_s corresponding to $\mathbf{m}_{ll}$ and $\mathbf{m}_{ss}$. Since the asymptotic theory in section 4 was developed for various conditions on n and r_i's, i.e., d_b and d_e, such theory should apply to a wide range of situations concerning $\mathbf{m}_{ll}$ and $\mathbf{m}_{ss}$.

For the general model (1.1) with either any number of hierarically nested random effects or with two balanced crossed random effects, we can find appropriate $\mathbf{m}_{ll}$ and $\mathbf{m}_{ss}$ for any covariance component $\boldsymbol{\Sigma}_{ll}$. As examples, the balanced two-way nested and crossed random effect structures are discussed here. For simplicity, we assume that the fixed effect $\mathbf{X}$ consists only of an overall mean.

First consider the balanced two-way nested model

$$\mathbf{y}_{ijk} = \mu + \mathbf{a}_i + \mathbf{b}_{ij} + \mathbf{e}_{ijk}, \quad \begin{aligned} i &= 1,2,\ldots,n_a, \\ j &= 1,2,\ldots,n_b, \\ k &= 1,2,\ldots,n_e, \end{aligned}$$

where the $p \times 1$ random vectors $\mathbf{a}_i, \mathbf{b}_{ij}$, and $\mathbf{e}_{ijk}$ are independent and normally distributed with zero mean and covariance matrices $\boldsymbol{\Sigma}_{aa}$, $\boldsymbol{\Sigma}_{bb}$, and $\boldsymbol{\Sigma}_{ee}$, respectively.

By fitting the $\mathbf{b}_{ij}$ effect after the $\mathbf{a}_i$ effect, we obtain the meann square matrices $\mathbf{m}_{aa}$, $\mathbf{m}_{bb}$, and $\mathbf{m}_{ee}$ with degrees of freedom $n_a - 1$, $n_a(n_b - 1)$, and $n_a n_b(n_e - 1)$, respectively. For testing the null that rank $\Sigma_{aa} \leq m$, we can use $\mathbf{m}_{aa}$ and $\mathbf{m}_{bb}$ and the corresponding T statistic, since the pairwise difference restriction (5.1) is satisfied with $\Sigma_{00} = n_e \Sigma_{bb} + \Sigma_{ee}$. For testing the null that rank $\Sigma_{bb} \leq m$, we use the test statistic T based on $\mathbf{m}_{bb}$ and $\mathbf{m}_{ee}$. Here also, the pairwise difference restriction is satisfied with $\Sigma_{00} = \Sigma_{ee}$. The asymptotic theory discussed in section 3 applies here with $n = n_a$, $r_i = n_b n_e$, $d_b = n_a - 1$, and $d_e = n_a(n_b - 1)$ for testing rank $\Sigma_{bb} \leq m$, and $n = n_a n_b$, $r_i = n_e$, $d_b = n_a(n_b - 1)$ and $d_e = n_a n_b(n_e - 1)$ for testing rank $\Sigma_{bb} \leq m$. Thus, we can develop a test statistic which can be used for a wide range of $(n_a,\ n_b,\ n_e)$ configurations.

For the second example, we consider the balanced two-way crossed random model

$$\mathbf{y}_{ijk} = \mu + \mathbf{a}_i + \mathbf{b}_j + (\mathbf{ab})_{ij} + \mathbf{e}_{ijk}, \quad \begin{aligned} i &= 1,2,\ldots,n_a, \\ j &= 1,2,\ldots,n_b, \\ k &= 1,2,\ldots,n_e, \end{aligned}$$

where the $p \times 1$ random vectors $\mathbf{a}_i$, $\mathbf{b}_j$, $(\mathbf{ab})_{ij}$, and $\mathbf{e}_{ijk}$ are independent and normally distributed with mean zero and covariance matrices Σ_{aa}, Σ_{bb}, Σ_{ab}, and Σ_{ee}, respectively. By fitting $(\mathbf{ab})_{ij}$ effect after $\mathbf{a}_i$ and $\mathbf{b}_j$ effects, we obtain the mean square matrices $\mathbf{m}_{aa}$, $\mathbf{m}_{bb}$, $\mathbf{m}_{ab}$, and $\mathbf{m}_{ee}$ with degrees of freedom $n_a - 1$, $n_b - 1$, $(n_a - 1)(n_b - 1)$, and $n_a n_b(n_e - 1)$, respectively. Note that for the balanced case the fitting order of $\mathbf{a}_i$ and $\mathbf{b}_j$ effects does not affect the results. If the model was unbalanced, we would fit the $\mathbf{a}_i$ effect before the $\mathbf{b}_j$ effect if we were interested in tesing rank Σ_{bb} and fit the $\mathbf{b}_j$ effect before the $\mathbf{a}_i$ effect if we were interested in testing rank Σ_{aa}. For testing the null that rank $\Sigma_{aa} \leq m$ for this balanced model, we use the statistic T based on $\mathbf{m}_{aa}$ and $\mathbf{m}_{ab}$ with their corresponding degrees of freedom, since the pairwise difference restriction (5.1) is satisfied with $\mathbf{\Sigma}_{00} = n_e\mathbf{\Sigma}_{ab} + \mathbf{\Sigma}_{ee}$. Similarly, for testing the null that rank $\mathbf{\Sigma}_{bb} \leq m$, we use $\mathbf{m}_{bb}$ and $\mathbf{m}_{ab}$. Finally, we use $\mathbf{m}_{ab}$ and $\mathbf{m}_{ee}$ for testing the null that rank $\mathbf{\Sigma}_{ab} \leq m$.

6. SIMULATION

To investigate the finite sample null and nonull properties of some special cases of the test procedures discussed in sections 2-4, a simulation study was conducted. The simulated model was the balanced one-way normal model (4.1) as defined in section 4 with p=4. The overall mean was set to zero. For the sample configuration (n, r), we considered three sets $(50, 5)$, $(50, 50)$, and $(5, 50)$, loosely corresponding to cases I, II, and III used to develop the asymptotic theory. We set the true rank m_0 of $\mathbf{\Sigma}_{bb}$ to be 2. For the true covariance components $\mathbf{\Sigma}_{bb}$ and $\mathbf{\Sigma}_{ee}$, we chose two parameterizations;

$$\text{(i) } \mathbf{\Sigma}_{bb}^{(\mathrm{i})} = \begin{pmatrix} 1 & 1 & .5 & .5 \\ 1 & 1 & .5 & .5 \\ .5 & .5 & 1 & 1 \\ .5 & .5 & 1 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_{ee}^{(\mathrm{i})} = \begin{pmatrix} 1 & .25 & .25 & .25 \\ .25 & 1 & .25 & .25 \\ .25 & .25 & 1 & .25 \\ .25 & .25 & .25 & 1 \end{pmatrix},$$

$$\text{(ii) } \mathbf{\Sigma}_{bb}^{(\mathrm{ii})} = \mathbf{\Sigma}_{bb}^{(\mathrm{i})}, \qquad \mathbf{\Sigma}_{ee}^{(\mathrm{ii})} = 30\mathbf{\Sigma}_{ee}^{(\mathrm{i})}.$$

Since the test statistic T is invariant under nonsingular transformations, an alternative (equivalent) parameterization is

$$\text{(i) } \mathbf{\Sigma}_{bb}^{(i)} = \operatorname{diag}\{1.714, 1.333, 0, 0\}, \qquad \mathbf{\Sigma}_{ee}^{(i)} = \mathbf{I}_4,$$

$$\text{(ii) } \mathbf{\Sigma}_{bb}^{(i)} = \operatorname{diag}\{0.057, 0.044, 0, 0\}, \qquad \mathbf{\Sigma}_{ee}^{(ii)} = \mathbf{I}_4.$$

For each of 6 combinations of the 3 sample configurations and the 2 parameterizations, 1000 samples were generated. One way to summarize the possible effects of the different parameterizations is to compute the roots $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ of the "expected determinantal equation"

$$\left|(r\Sigma_{bb} + \Sigma_{ee}) - \lambda\Sigma_{ee}\right| = 0.$$

The value of n has no effect on λ_i's. For all cases, $\lambda_3 = \lambda_4 = 1$. For λ_1 and λ_2,

$$\begin{aligned}
(\lambda_1, \lambda_2) &= (86.71,\ 67.67), && \text{i) with } r = 50, \\
&= (9.57,\ 7.67), && \text{i) with } r = 5, \\
&= (3.86,\ 3.22), && \text{ii) with } r = 50, \\
&= (1.29,\ 1.22), && \text{ii) with } r = 5.
\end{aligned}$$

Compared to parameterization i), parameterization ii) gives values of λ_1 and λ_2 closer to one, possibly making the detection of rank 2 structure more difficult. Also λ_1 and λ_2 become larger than one when r increases.

To study the null distribution as well as the power properties of the test procedures, we applied each procedure to test that rank $\Sigma_{bb} \leq m$ for $m = 0, 1, 2, 3$. Note that $m = 0$ case is the one-sided test for $\Sigma_{bb} = 0$, i.e., the test of the existence of the random effect. Because $m_0 = 2$, the hypothesis with $m = 2$ and 3 corresponds to the null, and the hypotheses with $m = 0$ and 1 are the nonull situations.

Note that for this model, $d_b = n-1$, $d_e = n(r-1)$, and the exact null distribution theory of section 5 applies. Thus, for T_W, T_B, T_H, and T_R in (2.7), the exact cut-off points can be obtained from Wilks lambda, Bartlett trace, Hotelling trace, and Roy largest root distributions. As discussed in section 4, the percentiles of these distributions have not been tabulated for the degrees of freedom combinations which may appear frequently in covariance component problems. Among the three combinations of (n, r) used here, only for the case with $(n, r) = (5, 50)$ were we able to find the exact cut-off points for T_W, T_B, T_H, and T_R. Using the 0.05 significance level, the simulation rejection frequencies for the exact test procedures are tabulated in Table 6.1. Note that the exact cut-off points correspond to the null distribution characterization given in the result of section 4. Because of the composite nature of the null hypothesis, the actual probability of rejection for our samples with $m = 2$ ranged from 0.004 to 0.045. We see that the powers testing for $m = 0, 1$ were smaller for parameterization ii) as conjectured by the above discussion on λ_i's. For our case, the four exact test procedures seem to have virtually equal null and nonnull powers.

Because the cut-off points for the exact test procedures are not available for the two other configurations of (n, r), and because the exact tests are not possible for more general models, we considered a number of asymptotic test procedures. For the form T given in (2.6) with Type 1 $g(\lambda, \gamma)$ fuction, we only considered $T_H = \sum_{i=m+1}^{p} \hat{\lambda}_i$ in (2.7) for simplicity, because the exact test procedures for $(n, r) = (5, 50)$ indicated the similarity among the four test statistics in (2.7). Theorems 3 and 5 showed that the asymptotic test procedures based on T_H for cases I, II, and III can be developed separately. In the theorems, c_1 for case I is r, $\frac{\partial g}{\partial \lambda} = 1$, $H_1 \sim N(0, 2(p-m))$, and $\sum_{i=1}^{p-m} g(\gamma_i, 0) \sim (n-1)\chi^2_{(p-m)(n-1-m)}$. Thus, the tests reject the null if $T_H > l_h$, h= I, II, III, where

$$l_{\mathrm{I}} = p - m + \sqrt{\frac{2r(p-m)}{(n-1)(r-1)}} z_\alpha,$$

$$l_{\mathrm{II}} = p - m + \sqrt{\frac{2(p-m)}{(n-1)}} z_\alpha,$$
$$l_{\mathrm{III}} = \frac{1}{n-1}\chi^2_\alpha,$$

z_α is the upper α point of $N(0,1)$, and χ^2_α is the upper α point of $\chi^2_{(p-m)(n-1-m)}$. Although these three tests were derived for the three different asymptotic situations, we applied these procedures to all three sample comfigurations (n, r). We denote these three test procedures by $T_{H,\,\mathrm{I}}$, $T_{H,\,\mathrm{II}}$, and $T_{H,\,\mathrm{III}}$. In Theorem 6, we derived a modification to cover cases I and II. Such a modification to $T_H = \sum_{i=m+1}^{p} \hat{\lambda}_i$ is to reject the null if

$$T_H > p - m + \sqrt{\frac{2(p-m)(nr-1)}{(n-1)n(r-1)}} z_\alpha.$$

We denote this test procedure, applied to all (n, r), by $T_{1,\,\mathrm{H}}$. Theorem 7 and the subsequent discussion gave a modified test using T_H^* in (3.3) which can be justified asymptotically for all cases I, II, and III. This test rejects the null if

$$T_H > (p-m)\left(1 - \sqrt{\frac{nr-1}{n(r-1)}}\right) + \frac{1}{n-1}\chi^2_\alpha.$$

We denote this procedure by T_H^*.

For T in (2.6) with Type 2 $g(\lambda, \delta)$ function, we considered T_L in (2.8). Note that T_L is the likelihood ratio test statistic for the model used for this simulation. By Theorems 4 and 5, for both cases I and II, the asymptotic cut-off point for $(n-1)T_L$ is B_α, where B_α is the upper α point of the distribution described and tabulated in Amemiya et al. (1990). By Theorem 6, this is also the modified test covering both cases I and II. We denote this test, applied to all (n, r) configurations, by T_L. The cut-off points for case III asymptotics in Theorem 5 and for the modified test covering all three cases are unavailable.

There is a possible test procedure for the two-sided testing problem that rank $\Sigma_{bb} = m$ against $\neq m$. Amemiya and Fuller (1984) developed a goodness-of-fit test procedure for a related problem. Their test applied to our problem can be considered to be the two-sided test in the sense of checking the goodness-of-fit of the model with rank m. This test rejects the null if

$$\begin{aligned} G &= \sum_{i=1}^{m} f(\hat{\lambda}_i, \frac{n-1}{n(r-1)}) I(\hat{\lambda}_i < 1) + \sum_{i=m+1}^{p} f(\hat{\lambda}_i, \frac{n-1}{n(r-1)}) \\ &> \frac{1}{n-1} \chi^2_{(p-m)(p-m+1)/2,\ \alpha}, \end{aligned}$$

where $f(\lambda, \delta)$ is as given in (11), and $\chi^2_{(p-m)(p-m+1)/2,\ \alpha}$, is the upper α point of χ^2 with $(p-m)(p-m+1)/2$ degrees of freedom. Note that G includes $\hat{\lambda}_i$ $i = 1, \ldots, m$ if $\hat{\lambda}_i < 1$ indicating rank possibly smaller than m. We included this test procedure denoted by G for our simulation.

These 7 asymptotic test procedures were applied to test for m=0, 1, 2, 3 in all 6 combinations of the sample configurations and parameterizations. Table 6.2 presents the simulated probabilities of rejection when the nominal significance level of 0.05 was used.

We first make general observations cocerning the effects of the different parameterizations (i) and (ii). Parameterization (ii) corresponds to the case that the errors (within-group variability) are very large compared to the random effect of interest. For this parameterization, all the test procedures tend to fail to reject the null very often, resulting in much smaller power to detect the alternative. For (ii), only the large sample case with $(n, r) = (50, 50)$ gives a reasonable power function for any of the tests. Testing the rank of a small random effect relative to the error seems difficult without large data. For parameterization (i), the power function of any test indicates a reasonable power to distinguish between the null and the alternative, regardless of the sample configuration and whether ot not the actual type I error is close to the nominal level.

As expected, the power is larger for testing $m = 0$ than $m = 1$, i.e., for $m_0 = m + 2$ than $m + 1$. As the theory indicates, for $m = 2$ and 3 corresponding to the null, i.e., $m = m_0$ and $m = m_0 - 1$, the maximum type I error occurs when $m = 2$.

For the two-sided test G, we see that the power is very small for $m_0 < m$ ($m = 3$) and that the power for $m_0 > m$ is often smaller than other tests. It seems that a two-sided rank problem can not be tested meaningfully. This is not very surprising, because the larger rank corresponds to the larger essential number of parameters. We recommend that the rank testing peroblem be formulated as a one-sided problem. We also note that the χ^2 null approximation for G does not work very well for small n.

For the likelihood ratio test T_L, we used the cut-off point justified asymptotically only for cases I and II, i.e., the cases loosely correspoinding to $(n, r) = (50, 5)$ and $(50, 50)$. This cut-off point seems to give the actual significance much smaller than the nominal level, producing relatively small power to detect the alternative compared to the test procedures related to T_H.

Considering the 5 test procedures based on T_H, we first notice the similarity between the performance of $T_{H,\mathrm{I}}$ and $T_{H,\mathrm{II}}$. Since $T_{1,H}$ is, in some sense, a compromize between $T_{H,\mathrm{I}}$ and $T_{H,\mathrm{II}}$, its performance is also similar to $T_{H,\mathrm{I}}$ and $T_{H,\mathrm{II}}$. Common features of these three tests are the very small type I error for parameterization (ii) and/or $n = 5$ and the corresponding small power for these cases. The procedure $T_{H,\mathrm{III}}$ uses the cut-off point justifiable only for small n and large r case. But, this test did not produce unreasonnable type I error for $(n, r) = (50, 5)$ and $(50, 50)$. This $T_{H,\mathrm{III}}$ has the common deficiency as $T_{H,\mathrm{I}}$, $T_{H,\mathrm{II}}$, and $T_{1,H}$, in that the power to detect a larger rank than m is very small for parameterization (ii) with small r. The test using T_H^* was derived so that the cut-off point is asymptotically justifiable for all cases I, II, and III. This test generally rejects the null more often than other tests. As a result, T_H^* is the only test with a reasonnable power property for parameterization (ii) and $r = 5$. But, for (i) with $r = 5$, the actual significance level exeeded the nominal level.

A possible conclusion is the recommendation of the practical use of T_H^* unless r is very small and the correct significance level is very important. Tabulation of the cut-off points for T_L under case III may be of interest. Such points can also be used for the modified test in Theorem 7, because T_L has a special structure $T_L^* = T_L$.

Table 6.1: Simulated rejection frequencies for the exact tests. $n = 5$, $r = 50$, $\alpha = 0.05$, $m_0 = 2$

Test	m	Parameterization (i)	Parameterization (ii)
T_W	0	1	0.704
	1	0.960	0.137
	2	0.038	0.004
	3	0.003	0.001
T_B	0	1	0.717
	1	0.962	0.145
	2	0.039	0.004
	3	0	0
T_H	0	1	0.731
	1	0.963	0.156
	2	0.045	0.005
	3	0	0
T_R	0	1	0.717
	1	0.962	0.145
	2	0.039	0.004
	3	0	0

Table 6.2: Simulated rejection frequencies for the asymptotic tests ($\alpha = 0.05$, $m_0 = 2$)

		Parameterization (i)			Parameterization (ii)		
		(n, r)					
Test	m	(50, 5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
T_L	0	1	1	1	0.220	1	0.570
	1	1	1	0.939	0.004	0.996	0.043
	2	0.023	0.024	.003	0	0.022	0
	3	0.001	0.001	0	0	0	0
G	0	1	1	1	0.147	1	0.775
	1	1	1	0.975	0.035	0.992	0.550
	2	0.050	0.057	0.614	0.110	0.062	0.694
	3	0.174	0.190	0.821	0.421	0.212	0.884
$T_{H,I}$	0	1	1	1	0.404	1	0.767
	1	1	1	0.952	0.020	0.992	0.072
	2	0.038	0.032	0.003	0	0.024	0
	3	0.002	0.001	0	0	0	0
$T_{H,II}$	0	1	1	1	0.463	1	0.767
	1	1	1	0.952	0.035	0.992	0.073
	2	0.048	0.032	0.003	0	0.025	0
	3	0.002	0.001	0	0	0	0
$T_{H,III}$	0	1	1	1	0.451	1	0.745
	1	1	1	0.963	0.047	0.993	0.158
	2	0.077	0.050	0.004	0.002	0.033	0.005
	3	0.004	0.001	0.004	0	0.002	0.001
$T_{1,H}$	0	1	1	1	0.46	1	0.767
	1	1	1	0.952	0.020	0.992	0.072
	2	0.038	0.032	0.003	0	0.024	0
	3	0.002	0.001	0	0	0	0
T_H^*	0	1	1	1	0.792	1	0.750
	1	1	1	0.963	0.224	0.993	0.159
	2	0.224	0.058	0.045	0.014	0.041	0.005
	3	0.014	0.001	0.004	0	0.003	0.001

7. DERIVATIONS

We present derivations of the results in section 3. Throughout this section, we assume that model (1.2) holds with assumptions (a), (b), and (c). To derive the limiting distributions of the roots in Theorems 1 and 2, we need to derive a nondegenerate limiting distribution of $(\mathbf{m}_{bb}, \mathbf{m}_{ee})$. First, we consider $\mathbf{m}_{ee}$.

Lemma 1. The matrices $\mathbf{m}_{bb}$ and $\mathbf{m}_{ee}$ are independent. For cases I, II, and III. $\sqrt{nr}(\mathbf{m}_{ee} - \mathbf{I}_p) \xrightarrow{L} \mathbf{V}$, where the $p \times p$ symmetric $\mathbf{V}$ consists of $p(p+1)/2$ independent normal random variables with mean zero and covariance $2\gamma^2$ for diagonal and γ^2 for off-diagonal elements, $\gamma^2 = c_1^{-1}$ for cases II and III, and $\gamma^2 = (c_1 - 1)^{-1}$ for case I with $r = 1$.

Proof. The independence follows from the fitting-constant definitions (2.2) and (2.3), and from the fact that $\mathbf{m}_{ee}$ is free of $\mathbf{b}_i$'s. Because $d_e \mathbf{m}_{ee} \sim W_p(\mathbf{I}_p, d_e)$, $d_e^{1/2}(\mathbf{m}_{ee} - \mathbf{I}_p)$ converges in distribution to a normal matrix as $d_e \to \infty$. By (b) and (c), $d_e = \sum_{i=1}^{n} r_i - n - k + l$ for fixed k and l, and thus

$$d_e n^{-1} r^{-1} \to \begin{cases} c_1 & \text{for case II and III,} \\ c_1 - 1 & \text{for case I with r = 1.} \end{cases}$$

Hence the result follows.

As noted in section 3, for cases II and III, different parts of $\mathbf{m}_{bb}$ have different rates of convergence. Thus, we need to use different normalizers for different parts. We write

$$\mathbf{m}_{bb} = \begin{pmatrix} \mathbf{m}_{bb11} & \mathbf{m}_{bb12} \\ \mathbf{m}_{bb21} & \mathbf{m}_{bb22} \end{pmatrix},$$

where $\mathbf{m}_{bb11}$ is $m_0 \times m_0$. Under (a), we can write $\mathbf{b}_i = (\mathbf{b}'_{1i}, \mathbf{0})'$ with probability one, where $m_0 \times 1$ $\mathbf{b}_{1i}$'s satisfy $\mathrm{Var}\{\mathbf{b}_{1i}\} = \mathbf{D}$ for $\mathbf{D}$ given in (a). We also write $\mathbf{e}_{ij} = (\mathbf{e}'_{1ij}, \mathbf{e}'_{2ij})'$ for $m_0 \times 1$ $\mathbf{e}_{1ij}$, $\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2)$ for $N \times m_0$ $\mathbf{E}_1$, and $\mathbf{B}_1 = (\mathbf{b}_{1i}, \mathbf{b}_{12}, \ldots, \mathbf{b}_{1n})'$. We first consider cases I and II, and obtain an expansion of $\mathbf{m}_{bb}$ for such cases.

Lemma 2. For cases I and II,

$$\begin{aligned}
\frac{1}{r}\mathbf{m}_{bb11} &= \frac{1}{nr}\sum_{i=1}^{n} r_i(\mathbf{b}_{1i} + \bar{\mathbf{e}}_{1i.})(\mathbf{b}_{1i} + \bar{\mathbf{e}}_{1i.})' + O_p(\frac{1}{n}), \\
\frac{1}{\sqrt{r}}\mathbf{m}_{bb12} &= \frac{1}{n\sqrt{r}}\sum_{i=1}^{n} r_i(\mathbf{b}_{1i} + \bar{\mathbf{e}}_{1i.})\bar{\mathbf{e}}'_{2i.} + O_p(\frac{1}{n}), \\
\mathbf{m}_{bb22} &= \frac{1}{n}\sum_{i=1}^{n} r_i\bar{\mathbf{e}}_{2i.}\bar{\mathbf{e}}'_{2i.} + O_p(\frac{1}{n}),
\end{aligned}$$

where r can be taken to be one for case I, and $\bar{\mathbf{e}}_{li.} = r_i^{-1}\sum_{j=1}^{r_i} \mathbf{e}_{lij}, l = 1, 2.$

Proof. Note that the weight matrix for $\mathbf{m}_{bb}$ in (2.2) can be expressed as

$$\mathbf{P}_{\mathbf{M_X Z}} = \mathbf{P_Z} + \mathbf{P}_{\mathbf{M_Z X}} - \mathbf{P_X},$$

where the projection matrix notations are defined in (2.1). Thus

$$\begin{aligned}
d_b\mathbf{m}_{bb11} &= (\mathbf{B}'_1\mathbf{Z}' + \mathbf{E}'_1)(\mathbf{P_Z} + \mathbf{P}_{\mathbf{M_Z X}} - \mathbf{P_X})(\mathbf{ZB}_1 + \mathbf{E}_1) \\
&= (\mathbf{B}'_1\mathbf{Z}' + \mathbf{E}'_1)\mathbf{P_Z}(\mathbf{ZB}_1 + \mathbf{E}_1) + \mathbf{E}'_1\mathbf{P}_{\mathbf{M_Z X}}\mathbf{E}_1 \\
&- (\mathbf{B}'_1\mathbf{Z}' + \mathbf{E}'_1)\mathbf{P_X}(\mathbf{ZB}_1 + \mathbf{E}_1)
\end{aligned}$$

$$= \sum_{i=1}^{n} r_i(\mathbf{b}_{1i} + \bar{\mathbf{e}}_{1i.})(\mathbf{b}_{1i} + \bar{\mathbf{e}}_{1i.})' + \mathbf{E}_1'\mathbf{P}_{M_Z\mathbf{X}}\mathbf{E}_1 - \mathbf{E}_1'\mathbf{P}_\mathbf{X}\mathbf{E}_1$$
$$- \mathbf{F}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}_1 - \mathbf{F}_1'\mathbf{F}_2 - \mathbf{F}_2'\mathbf{F}_1, \quad (7.1)$$

where $\mathbf{F}_1 = \sum_{i=1}^{n} \bar{\mathbf{x}}_{i.}'\mathbf{b}_{1i}$ and $\mathbf{F}_2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{E}_1$. By assumptions (a) and (b), the second and third terms in (7.1) are Wishart matrices with degrees of freedom $(k-l)$ and k respectively, and both are $O_P(1)$. Both $\mathbf{F}_1$ and $\mathbf{F}_2$ have zero mean, and the variances of the elements are $O(nr^2)$ and $O(n^{-1}r^{-1})$, respectively, by assumption (c). Hence, the fourth terms in (7.1) is $O_P(r)$, and the last two terms are $O_P(r^{1/2})$. The result for $\mathbf{m}_{bb11}$ follows since $d_b = n - l$ with fixed l. For $\mathbf{m}_{bb12}$,

$$d_b\mathbf{m}_{bb12} = (\mathbf{B}_1'\mathbf{Z}' + \mathbf{E}_1')\mathbf{P}_\mathbf{Z}\mathbf{E}_2 + \mathbf{E}_1'\mathbf{P}_{\mathbf{M_Z X}}\mathbf{E}_2 - \mathbf{E}_1'\mathbf{P}_\mathbf{X}\mathbf{E}_2 - \mathbf{F}_1'\mathbf{F}_3,$$

where $\mathbf{F}_3 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}_2 = O_P(n^{-1/2}r^{-1/2})$, and the second and third terms are parts of Wishart matrices with fixed degrees of freedom. The result follows, because the first term is $\sum_{i=1}^{n} r_i(\mathbf{b}_{1i} + \bar{\mathbf{e}}_{1i.})'\bar{\mathbf{e}}_{2i.}$. The result for $\mathbf{m}_{bb22}$ is immediate, since $d_b\mathbf{m}_{bb22} = \mathbf{E}_2'\mathbf{P}_\mathbf{Z}\mathbf{E}_2 + \mathbf{E}_2'\mathbf{P}_{\mathbf{M_Z X}}\mathbf{E}_2 - \mathbf{E}_2'P_\mathbf{X}\mathbf{E}_2$.

Now, we derive the nondegenerate limiting distribution of $\mathbf{m}_{bb}$, for cases I and II.

Lemma 3. For cases I and II,

$$\sqrt{n}\begin{pmatrix} \frac{1}{\sqrt{r}}\mathbf{I}_{m_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-m_0} \end{pmatrix}\left[\mathbf{m}_{bb} - \frac{1}{n}\sum_{i=1}^{n} r_i\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} - \mathbf{I}_p\right]\begin{pmatrix} \frac{1}{\sqrt{r}}\mathbf{I}_{m_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-m_0} \end{pmatrix}$$

$$\xrightarrow{L} \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix},$$

where the elements of $\mathbf{U}$ are jointly normally distributed with mean zero, the functionally independent elements of $(p-m_0)\times(p-m_0)$ symmetric $\mathbf{U}_{22}$ are independent with variances 2 for diagonal and 1 for off-diagonal elements.

Proof. Let $\mathbf{w}_i = (\mathbf{w}_{1i}', \mathbf{w}_{2i}')' = \sqrt{r_i}(\bar{\mathbf{e}}_{1i.}', \bar{\mathbf{e}}_{2i.}')'$. Then, $\mathbf{w}_i$'s are $N_p(\mathbf{0}, \mathbf{I}_p)$ random vectors. By Lemma 2, we can write

$$
\begin{aligned}
\frac{1}{r}[\mathbf{m}_{bb11} - \frac{1}{n}\sum_{i=1}^{n} r_i\mathbf{D} - \mathbf{I}_{m_0}] &= \frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{r}(\mathbf{b}_{1i}\mathbf{b}_{1i}' - \mathbf{D}) \\
&+ \sum_{i=1}^{n}\frac{\sqrt{r_i}}{r}(\mathbf{b}_{1i}\mathbf{w}_{1i}' + \mathbf{w}_{1i}\mathbf{b}_{1i}') \\
&+ \frac{1}{r}\sum_{i=1}^{n}(\mathbf{w}_{1i}\mathbf{w}_{1i}' - \mathbf{I}_{m_0}) + O_p(\frac{1}{n}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{r}(\mathbf{b}_{1i}\mathbf{b}_{1i}' - \mathbf{D}) \\
&+ O_p[max(n^{-1/2}r^{-1/2}, n^{-1})], \text{ for case II,}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\sqrt{r}}\mathbf{m}_{bb12} &= \frac{1}{n}\sum_{i=1}^{n}[\sqrt{\frac{r_i}{r}}\mathbf{b}_{1i}\mathbf{w}_{2i}' + \frac{1}{\sqrt{r}}\mathbf{w}_{1i}\mathbf{w}_{2i}'] + O_p(\frac{1}{n}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{r_i}{r}}\mathbf{b}_{1i}\mathbf{w}_{2i}' + O_p[max(n^{-1/2}r^{-1/2}, n^{-1})], \text{ for case II,}
\end{aligned}
$$

$$
\mathbf{m}_{bb22} - \mathbf{I}_{p-m_0} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{w}_{2i}\mathbf{w}_{2i}' - \mathbf{I}_{p-m_0}) + O_p(\frac{1}{n}).
$$

By assumption (a), the elements of $(\mathbf{b}_{1i}\mathbf{b}_{1i}' - \mathbf{D}), \mathbf{b}_{1i}\mathbf{w}_{1i}', \mathbf{w}_{1i}\mathbf{w}_{1i}', \mathbf{b}_{1i}\mathbf{w}_{2i}', \mathbf{w}_{1i}\mathbf{w}_{2i}'$, $(\mathbf{w}_{2i}\mathbf{w}_{2i}' - \mathbf{I}_{p-m_0})$ are jointly, independently, and identically distributed with zero mean and finite second moments. Also, by assumption (c),

$$
\frac{1}{n}\sum_{i=1}^{n}(\frac{r_i}{r})^2 \rightarrow c_2,
$$

$$\frac{1}{n}\sum_{i=1}^{n}(\frac{\sqrt{r_i}}{r})^2 \rightarrow \frac{c_2}{r}, \text{ for case I,}$$
$$\frac{1}{n}\sum_{i=1}^{n}(\frac{\sqrt{r_i}}{\sqrt{r}})^2 \rightarrow c_1.$$

Thus, the limiting normality follows by a standard limit theorem [see, e.g., Lemma 1 in Amemiya and Fuller (1984)]. The variances for the elements of $\mathbf{U}_{22}$ are consequences of the distribution of $\mathbf{w}_{2i}$'s.

Lemmas 1 and 3 can be used to derive Theorem 1.

Proof of Theorem 1. We apply a general theorem on the limiting distributions of the roots given in Theorem 1 (and its discussion) of Remadi and Amemiya (1992). By Lemmas 1 and 3, this theorem applies here with $a_n = \sqrt{n}$, $b_n = \sqrt{nr}$, $c_n = \sqrt{r}$, $\gamma = r^{-1/2}$ for case I and $= 0$ for case II, and $\mathbf{D}_2(n) = \mathbf{I}_{p-m_0}$. We take $r = 1$ for case I. Thus $\sqrt{n}(\hat{\lambda}_i - 1)$, $i = m_0+1, m_0+2, \ldots, p$, jointly converge in distribution to the roots of $\mathbf{U}_{22} - \mathbf{V}_{22}$ for case I and of $\mathbf{U}_{22}$ for case II, where $\mathbf{U}_{22}$ is given in Lemma 3, and $\mathbf{V}_{22}$ is the $(p-m_0)\times(p-m_0)$ lower right-hand corner of $\mathbf{V}$ in Lemma 1. The result follows from the independence of $\mathbf{U}_{22}$ and $\mathbf{V}_{22}$ and the distribution of $\mathbf{U}_{22}$ and $\mathbf{V}_{22}$ given in Lemmas 1 and 3.

To derive Theorem 2 for case III, we use a conditional argument by first conditionning on $\mathbf{b}_i, i = 1, 2, \ldots, n$, where n is fixed. We first derive a conditional expansion for $\mathbf{m}_{bb}$.

Lemma 4. For case III, under the additional assumption (d), conditionally on $\mathbf{b}_i's$,

$$\frac{1}{r}\mathbf{m}_{bb11} = \frac{1}{d_b r}\mathbf{B}_1'\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{Z}\mathbf{B}_1 + O_p(\frac{1}{\sqrt{r}}),$$

$$\begin{aligned}\frac{1}{\sqrt{r}}\mathbf{m}_{bb12} &= \frac{1}{d_b\sqrt{r}}\mathbf{B}_1'\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{E}_2 + O_p(\frac{1}{\sqrt{r}}),\\ \mathbf{m}_{bb22} &= \frac{1}{d_b}\mathbf{E}_2'\mathbf{P}_{\mathbf{M}_\mathbf{X}\mathbf{Z}}\mathbf{E}_2,\end{aligned}$$

where $\mathbf{B}_1$ and $\mathbf{E}_2$ are difined above Lemma 2.

Proof. Note that

$$\begin{aligned}d_b\mathbf{m}_{bb11} &= (\mathbf{B}_1'\mathbf{Z}' + \mathbf{E}_1')\mathbf{P}_{\mathbf{M}_\mathbf{X}\mathbf{Z}}(\mathbf{Z}\mathbf{B}_1 + \mathbf{E}_1) \qquad (7.2)\\ &= \mathbf{B}_1'\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{Z}\mathbf{B}_1 + \mathbf{B}_1'\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{E}_1 + \mathbf{E}_1'\mathbf{M}_\mathbf{X}\mathbf{Z}\mathbf{B}_1 + \mathbf{E}_1'\mathbf{P}_{\mathbf{M}_\mathbf{X}\mathbf{Z}}\mathbf{E}_1.\end{aligned}$$

The variances of the elements of the $(n \times p)$ $\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{E}_1$ have the order of $\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{Z}$, i.e., $O(r)$ by assumption (d), and the last term in (7.2), being a Wishart matrix with $(n-l)$ degrees of freedom is $O_p(1)$. The result for $\mathbf{M}_{bb11}$ follows since $d_b = n-l$ is fixed for case III. For $\mathbf{m}_{bb12}$, the term $\mathbf{E}_1'\mathbf{P}_{\mathbf{M}_X\mathbf{Z}}\mathbf{E}_2$ is $O_p(1)$. Thus, the result follows.

We next derive the conditional limiting distribution of $\mathbf{m}_{bb}$ for case III.

Lemma 5. For case III with assumption (d), conditionally on $\mathbf{b}_i$'s,

$$\begin{pmatrix} \frac{1}{\sqrt{r}}\mathbf{I}_{m_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-m_0} \end{pmatrix} \mathbf{m}_{bb} \begin{pmatrix} \frac{1}{\sqrt{r}}\mathbf{I}_{m_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-m_0} \end{pmatrix} \xrightarrow{L} \mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix},$$

where $d_b\mathbf{R}$ has the noncentral Wishart distribution with degrees of freedom $n-l$, covariance matrix block diag$\{\mathbf{B}_1'\mathbf{K}\mathbf{B}_1,\ \mathbf{0}\}$.

Proof. By Lemma 4, the leading term of the quantity considered in this Lemma can be written as

$$\mathbf{A}_r = \frac{1}{d_b}\left(\frac{1}{\sqrt{r}}\mathbf{Z}\mathbf{B}_1,\ \mathbf{E}_2\right)'\mathbf{P}_{\mathbf{M}_\mathbf{X}\mathbf{Z}}\left(\frac{1}{\sqrt{r}}\mathbf{Z}\mathbf{B}_1,\ \mathbf{E}_2\right).$$

Thus, given $\mathbf{B}_1$, $d_b\mathbf{R} \sim W_p($ block diag $\{\mathbf{0},\ \mathbf{I}_{p-m_0}\}$, $n-l$, block diag $\{\mathbf{\Phi}_r,\ \mathbf{0}\})$, where $\mathbf{\Phi}_r = \frac{1}{r}\mathbf{B}_1'\mathbf{Z}'\mathbf{M}_\mathbf{X}\mathbf{Z}\mathbf{B}_1$. Let $m_0 \times m_0$ $\mathbf{L}_r$ be the cholesky decomposition of $\mathbf{\Phi}_r$

so that $\Phi_r = \mathbf{L}_r\mathbf{L}_r'$. Since the elements of $\mathbf{L}_r$ are continuous functions of Φ_r, and since $\Phi_r \to \mathbf{B}_1'\mathbf{K}\mathbf{B}_1$ as $r \to \infty$, it follows that

$$\mathbf{L}_r \to \mathbf{L}_0, \tag{7.3}$$

with $\mathbf{L}_0$ satisfying $\mathbf{B}_1\mathbf{K}\mathbf{B}_1 = \mathbf{L}_0\mathbf{L}_0'$. The distribution of $d_b\mathbf{A}_r$ is the same as that of a $\sum_{t=1}^{n-l} \mathbf{a}_t\mathbf{a}_t'$, where $\mathbf{a}_t = (\mathbf{a}_{1t}', \mathbf{a}_{2t}')'$, $\mathbf{a}_{1t}$ is the t-th column of $\mathbf{L}_r$ for $t = 1, 2, \ldots, m_0$, and is 0 for $t = m_0 + 1, m_0 + 2, \ldots, n - l$, and $\mathbf{a}_{2t}, t = 1, 2, \ldots, n - l$, are independent $N_{p-m_0}(\mathbf{0}, \mathbf{I}_{p-m_0})$ random vectors. Thus, by (7.3), as $r \to \infty$ with fixed n, $\sum_{t=1}^{n-l} \mathbf{a}_t\mathbf{a}_t'$ converges in distribution to $d_b\mathbf{R}$ as given in the lemma, and so does $d_b\mathbf{A}_r$.

Using this Lemma on the conditional limiting distribution, we now give a proof of Theorem 2.

Proof of Theorem 2. First, we condition on n (fixed) $\mathbf{b}_i$'s, such that $\mathbf{B}_1'\mathbf{K}\mathbf{B}_1$ is nonsingular. By assumption (d), such $\mathbf{b}_i$'s constitute a set of probability one. Because $\mathbf{m}_{ee}$ is free of the $\mathbf{b}_i$'s, and because $\mathbf{e}_{ij}$'s are independent of the $\mathbf{b}_i$'s, the limiting distributions in Lemmas 1 and 5 specify the joint conditional limiting distribution of $\mathbf{m}_{ee}$ and $\mathbf{m}_{bb}$ given $\mathbf{b}_i$'s. Thus, Theorem 2 of Remadi and Amemiya (1992) applies to this conditional sequence, and we obtain that conditionally, $\hat{\lambda}_i$, $i = m_0 + 1, m_0 + 2, \ldots, p$ converge in distribution to the ordered roots of $\mathbf{R}_{22.1} = \mathbf{R}_{22} - \mathbf{R}_{21}\mathbf{R}_{11}^{-1}\mathbf{R}_{12}$, where $\mathbf{R}$ is defined in Lemma 5. Note that $\mathbf{R}_{11} = \mathbf{B}_1'\mathbf{K}\mathbf{B}_1$ is nonsingular for almost all $\mathbf{B}_1$, because $\mathbf{b}_i$'s are independent and identically distributed with nonsingular covariance matrix, and rank $\mathbf{K} \geq p$ by assumption (d). See, Theorem 5 of Das Gupta (1971). By the distribution of $d_b\mathbf{R}$ given in Lemma 5 and the standard regression argument, $d_b\mathbf{R}_{22.1}$ has the Wishart distribution with covariance matrix $\mathbf{I}_{p-m_0}$ and degrees of freedom $n-l-m_0$. This conditional limiting distribution of $\hat{\lambda}_i$'s is free of the conditioned values of $\mathbf{b}_i$'s with probability one. Since the convergence of the bounded sequence of conditional distribution functions and the expectation with respect to $\mathbf{b}_i$'s can be interchanged by the Lebesgue convergence theorem, it follows that the conditional limiting distribution of $\hat{\lambda}_i$'s, $i = m_0+1, m_0+2, \ldots, p$, is also the unconditional limiting distribution.

Proof of Theorem 3. For Type 1 $g(\lambda, \delta)$, we have for each $\delta \geq 0$

$$g(\lambda,\delta) = g(1,\delta) + \frac{\partial g}{\partial \lambda}(\lambda^*, \delta)(\lambda - 1),$$

where λ^* is between 1 and λ. Thus, the results for cases I and II follow from Theorem 1, the continuity of $\frac{\partial g}{\partial \lambda}(\lambda, \delta)$, and the fact that

$$\begin{aligned} \frac{d_b}{d_e} &\rightarrow \frac{1}{c_1 - 1}, \text{ for case I,} \\ &\rightarrow 0, \text{ for cases II and III.} \end{aligned} \tag{7.4}$$

The result for case III is a consequence of Theorem 2, the continuity of $g(\lambda, \delta)$, and (7.4).

To prove Theorem 4, we need the following lemma because of the form of Type 2 $g(\lambda, \delta) = f(\lambda, \delta) I(\lambda > 1)$.

Lemma 6. If a sequence of random variables X_n and a sequence of nonnegative functions $g_n(.)$ satisfy $[X_n, g_n(X_n)] \xrightarrow{L} [X, g(X)]$, where the limiting distribution is continuous, then for every x,

$$g_n(X_n) I(X_n > x) \xrightarrow{L} g(X) I(X > x).$$

Proof. The result is immediate, because for every $y > 0$

$$\begin{aligned} P\{g_n(X_n)I(X_n > x) > y\} &= P\{g_n(X_n) > y \text{ and} X_n > x\} \\ &\to P\{g(X) > y \text{ and} X > x\} \\ &= P\{g(X)I(X > x) > y\}. \end{aligned}$$

Proof of Theorem 4. For Type 2 function $g(\lambda, \delta) = f(\lambda, \delta)I(\lambda > 1)$, we have for every $\delta \geq 0$,

$$f(\lambda, \delta) = \frac{\partial^2 f}{\partial \lambda^2}(\lambda^*, \delta)(\lambda - 1)^2,$$

where λ^* is between 1 and λ. Thus, Theorem 1, the continuity of $\frac{\partial^2 f}{\partial \lambda^2}(\lambda, \delta)$, (7.4), and Lemma 6 give the results for cases I and II. Note that Lemma 6 is applied to $X_n = \sqrt{n}(\hat{\lambda}_i - 1)$ and $x = 0$, and the limiting distributions as given in Theorem 1 are continuous. The result for case III is a consequence of Theorem 2, Lemma 6 (with $x = 1$), the continuity of $f(\lambda, \delta)$, and (7.4).

The following lemma is needed for deriving Theorem 5.

Lemma 7. For any symmetric matrix $\mathbf{C}$, let $\nu_i(\mathbf{C})$ be the $i-th$ largest eigenvalue. Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

be a $n \times n$ symmetric matrix, where $\mathbf{A}_{11}$ is $k \times k$ positive definite matrix. Define $\mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. Then, for $i = 1, 2, \ldots, n-k$

$$\begin{aligned} \nu_i(\mathbf{A}_{22}) &\geq \nu_{k+i}(\mathbf{A}), \\ \nu_i(\mathbf{A}_{22.1}) &\geq \nu_{k+i}(\mathbf{A}), \end{aligned}$$

Proof. The first result follows from the Poincare separation Theorem [See, e.g., Rao (1973, p. 64)] by noting that $\mathbf{A}_{22} = (0, \mathbf{I}_{n-k})\mathbf{A}(0, \mathbf{I}_{n-k})'$. For the second result, let

$$\mathbf{B} = \begin{pmatrix} -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{I}_{n-k} \end{pmatrix} \mathbf{C}^{-1/2},$$

where $\mathbf{C} = \mathbf{I}_{n-k} + \mathbf{A}_{21}\mathbf{A}_{11}^{-2}\mathbf{A}_{12}$ is positive definite. Thus, $\mathbf{B}'\mathbf{B} = \mathbf{I}_{n-k}$, and the Poincare separation Theorem gives

$$\nu_i(\mathbf{B}'\mathbf{A}\mathbf{B}) \geq \nu_{k+i}(\mathbf{A}).$$

Note that $\mathbf{B}'\mathbf{A}\mathbf{B} = \mathbf{C}^{-1/2}\mathbf{A}_{22.1}\mathbf{C}^{-1/2}$. By Theorem 2.2 (with $\mathbf{B} = \mathbf{C}$, $\mathbf{C} = \mathbf{I}, j = i, k = 1$) of Anderson and Das Gupta (1963), $\nu_i(\mathbf{C}^{-1/2}\mathbf{A}_{22.1}\mathbf{C}^{-1/2}) \leq \nu_i(\mathbf{A}_{22.1})\nu_1(\mathbf{C}^{-1})$. The second result follows because $\nu_1(\mathbf{C}^{-1}) \leq 1$.

Proof of Theorem 5. Fix an $m_0 = 0, 1, \ldots, m-1$. The quantities $M_l(h,\ m_0)$, $l = 1, 2,\ h = \mathrm{I}, \mathrm{II}$, are functions of γ_i, $i = m - m_0 + 1, m - m_0 + 2, \ldots, p - m_0$, the $p - m$ smaller roots of $(p - m_0) \times (p - m_0)$ $\mathbf{G}(m_0)$, where $\mathbf{G}(m_0)$ is as given in Theorem 1 for cases I and II and in Theorem 2 for case III. We write these γ_i's as γ_{m-m_0+j}, $j = 1, 2, \ldots, p - m$. Let

$$\mathbf{G}(m_0) = \begin{pmatrix} \mathbf{G}_{11}(m_0) & \mathbf{G}_{12}(m_0) \\ \mathbf{G}_{21}(m_0) & \mathbf{G}_{22}(m_0) \end{pmatrix},$$

$\mathbf{G}_{22.1}(m_0) = \mathbf{G}_{22}(m_0) - \mathbf{G}_{21}(m_0)\mathbf{G}_{11}^{-1}(m_0)\mathbf{G}_{12}(m_0)$, where $\mathbf{G}_{11}(m_0)$ is $(m - m_0) \times (m - m_0)$ and $\mathbf{G}_{22}(m_0)$ is $(p - m) \times (p - m)$. Then, by Lemma 7, with probability one, for $j = 1, 2, \ldots, p - m$

$$\nu_j(\mathbf{G}_{22}(m_0)) \geq \gamma_{m-m_0+j} \tag{7.5}$$

$$\nu_j(\mathbf{G}_{22.1}(m_0)) \geq \gamma_{m-m_0+j} \tag{7.6}$$

where $\nu_j(.)$ is defined in Lemma 7. Note that every $M_l(h,\ m_0)$ has the form

$$\sum_{j=1}^{a} \Gamma(\gamma_{m-m_0+j})$$

for some nondecreasing function $\Gamma(.)$. For cases I and II and for every x, by (7.5),

$$\begin{aligned} P\left\{\sum_{j=1}^{a} \Gamma(\gamma_{m-m_0+j}) > x\right\} &\leq P\left\{\sum_{j=1}^{a} \Gamma[\nu_j(\mathbf{G}_{22}(m_0))] > x\right\} \\ &= P\left\{\sum_{j=1}^{a} \Gamma[\nu_j(\mathbf{G}(m)] > x\right\}, \end{aligned}$$

because the $(p-m)\times(p-m)$ normal matrix $\mathbf{G}_{22}(m_0)$ has the same distribution as $\mathbf{G}(m)$ as given in Theorem 1. For case III and any x, by (7.6),

$$\begin{aligned} P\left\{\sum_{j=1}^{a} \Gamma(\gamma_{m-m_0+j}) > x\right\} &\leq P\left\{\sum_{j=1}^{a} \Gamma[\nu_j(\mathbf{G}_{22.1}(m_0))] > x\right\} \\ &= P\left\{\sum_{j=1}^{a} \Gamma[\nu_j(\mathbf{G}(m)] > x\right\}, \end{aligned}$$

because the $(p-m)\times(p-m)$ $\mathbf{G}_{22.1}(m_0)$ is a Wishart matrix with degrees of freedom $(n-l-m_0)-(m-m_0) = n-l-m$, the same distribution as $\mathbf{G}(m)$ as given in Theorem 2. These hold for any $m_0 = 0, 1, \ldots, m-1$. The distribution of $K(l, h)$ follows by Theorems 3 and 4 with $m_0 = m$ and by the distribution of the roots as given in Theorems 1 and 2 with $m_0 = m$.

Proof of Theorem 6. This result follows by Theorems 3, 4, and 5 and by noting

$$\frac{d_b}{d_e} \to \begin{cases} (c_1-1)^{-1}, & \text{case I with } r=1, \\ 0, & \text{case II,} \end{cases}$$

$$\frac{d_e}{d_b+d_e} \to \begin{cases} c_1^{-1}(c_1-1), & \text{case I with } r=1 \\ 1, & \text{case II.} \end{cases}$$

For the proof of Theorem 7, we need the following Lemma.

Lemma 8. Let $x_{n,\alpha}$ and x_α be the upper α point $(0 < \alpha < 1)$ of X_n and X, $n = 1, 2, \ldots$. If $X_n \xrightarrow{L} X$ as $n \to \infty$, and if the distribution function $F(x)$ of X is continuous, then $x_{n,\alpha} \to x_\alpha$ as $n \to \infty$.

Proof. Let $F_n(x)$ be the distribution function of X_n. Since $F(x)$ is continuous, $F_n(x) \to F(x)$ as $n \to \infty$ uniformly for all x [Polya's Theorem, See, e.g., Rao (1973, p. 120)]. Hence, it follows that, for any sequence y_n with $y_n \to y$, $F_n(y_n) \to F(y)$. Because $X_n \xrightarrow{L} X$, X_n is tight [See, e.g., Billingsley (1986, p. 346)]. Thus, for each α $(0 < \alpha < 1)$, the sequence $\{x_{n,\alpha}\}$ is bounded. Hence, every subsequence has a converging subsubsequence $x_{n_l,\alpha} \to x\{n_l\}$ as $l \to \infty$, where the limit $x\{n_l\}$ depends on such a subsubsequence. By the above result on $y_n \to y$, as $l \to \infty$

$$\alpha = F_{X_{n_l}}(x_{n_l,\alpha}) \to F_X(x\{n_l\}),$$

But, by the continuity of $F_X(x)$, $x\{n_l\} = x_\alpha$ for all such converging subsequences. Thus, $x_{n,\alpha} \to x_\alpha$.

Proof of Theorem 7. Consider the random variables,

$$\begin{aligned} K(1,\mathrm{III}) &= \sum_{i=1}^{a} g(\gamma_i, 0)\,, \\ K(2,\mathrm{III}) &= \sum_{i=1}^{a} f(\gamma_i, 0) I(\gamma_i > 1) \end{aligned} \tag{7.7}$$

as given in Theorem 5, where γ_i's are the roots of $\mathbf{G}(m)$ and $d_b\mathbf{G}(m) \sim W_{p-m}(I_{p-m}, n-l-m)$ as given in Theorem 2. We consider these as indexed by n, since the distribution of γ_i's depend on n. Note that, as $n \to \infty$,

$$\sqrt{n}[G(m) - I_{p-m}] \xrightarrow{L} F_0,$$

where F_0 is as specified in Theorem 5. Thus, by a standard result on the limiting distribution of the roots [See, e.g. Anderson (1984, p. 542)], for γ_i's in (7.7), as $n \to \infty$,

$$\sqrt{n}(\gamma_1 - 1, \ldots, \gamma_{p-m} - 1) \xrightarrow{L} (\eta_1, \ldots, \eta_{p-m}),$$

where the η_i's are as given in Theorem 5. Using Lemma 6, we obtain that as $n \to \infty$

$$\begin{aligned} \sqrt{n}[K(1, \mathrm{III}) - ag(1,0)] &\xrightarrow{L} \frac{\partial g}{\partial \lambda}(1,0)H_1, \\ nK(2, \mathrm{III}) &\xrightarrow{L} \frac{\partial^2 f}{\partial \lambda^2}(1,0)H_2, \end{aligned}$$

where H_1 and H_2 are given in (3.2). Hence, by Lemma 8, as $n \to \infty$

$$\begin{aligned} \sqrt{n}[k_{1\alpha} - ag(1,0)] &\to \frac{\partial g}{\partial \lambda}(1,0)H_{1\alpha}, \\ nk_{2\alpha} &\to \frac{\partial^2 f}{\partial \lambda^2}(1,0)H_{2,\alpha}. \end{aligned}$$

Therefore, for cases I and II,

$$\begin{aligned} \sqrt{n}\left\{\frac{\partial g(1,0)}{\partial \lambda}\right\}^{-1}(T_1^* - k_{1\alpha}) &= \sqrt{n}\sqrt{\frac{d_e}{d_e + d_b}}\left\{\frac{\partial g(1, \frac{d_b}{d_e})}{\partial \lambda}\right\}^{-1}[T - ag(1, \frac{d_b}{d_e})] \\ &\quad -\sqrt{n}\left\{\frac{\partial g(1,0)}{\partial \lambda}\right\}^{-1}[k_{1\alpha} - ag(1,0)] \\ &\xrightarrow{L} H_1^* - H_{1,\alpha}, \end{aligned}$$

and

$$\begin{aligned} n\left\{\frac{\partial^2 f(1,0)}{\partial \lambda^2}\right\}^{-1}(T_2^* - k_{2\alpha}) &= n\frac{d_e}{d_e + d_b}\left\{\frac{\partial^2 f(1, \frac{d_b}{d_e})}{\partial \lambda^2}\right\}^{-1} T \\ &\quad -n\left\{\frac{\partial^2 f(1,0)}{\partial \lambda^2}\right\}^{-1} k_{2\alpha} \\ &\xrightarrow{L} H_2^* - H_{2,\alpha}, \end{aligned}$$

where, for $l = 1, 2$, $H_l^* = H_l$ if $m = m_0$, and H_l^* is stochastically smaller than H_l if $m_0 < m$. Thus, for cases I and II, $P\left\{T_l^* - k_{l\alpha} > 0\right\} \to P\left\{H_l^* > H_{l,\alpha}\right\}$, $l = 1, 2$. For case III, $\frac{d_b}{d_e} \to 0$ and $\frac{d_e}{d_b + d_e} \to 1$. Hence, for case III, $T_l^* - T \xrightarrow{P} 0$, $l = 1, 2$, and the result follows.

8. CONCLUSION

For a broad class of multivariate mixed effect models, a class of test procedures for teting the rank of a covariance component was introduced. Asymptotic properties of such procedures were derived under various conditions. Approximate test procedures which can be used for a wide range of applications were derived. For a simple special case, the test procedures with exact significance level were discussed. A simulation study supported the usefulness of the approximate test procedures.

This paper provided a uniform approach to test development and asymptotic theory in the rank testing problem. Statistical procedures useful for a large class of applied problems were presented.

REFERENCES CITED

Amemiya, Y. (1985), What should be done when an estimated between-group covariance matrix is not nonnegative definite. The American Statistician 39, 2, 112-117.

Amemiya, Y., T.W. Anderson and P.A.W. Lewis (1990). Percentage points for a test of rank in multivariate components of variance. Biometrika 77, 637-641.

Amemiya, Y. and W.A. Fuller (1984). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. Ann. Statist. 12, 497-509.

Anderson, B., T.W. Anderson and I. Olkin (1986). Maximum liklihood estimators and liklihood ratio criteria in multivariate components of variance. Ann. Statist. 14, 405-417.

Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis. (Wiley, New York, 2nd ed.).

Anderson, T.W. (1989). The asymptotic distribution of the liklihood ratio criterion for testing rank in multivariate components of variance. J. Multivariate Anal. 30, 72-79.

Anderson, T.W. and S. Das Gupta (1963). Some inequalities on characteristic roots of matrices. Biometrika 50, 522-523.

Anderson, T.W. and S. Das Gupta (1964). A monotonicity property of the power function of some tests of the equality of two covariance matrices. Ann. Math. Stat. 35, 1059-1063.

Anderson, T.W. and Y. Amemiya (1991). Testing dimensionality in the multivariate analysis of variance. Statist. Probab. Lett. 12, 445-463.

Billingsley, P. (1986) Probability and measure. Wiley, New york.

Das Gupta, S. (1971). Nonsingularity of the sample covariance matrix. Sankhya Ser A, 33, 475-478.

Fuller, W.A and G.E. Battese (1973). Transformations for estimation of linear models with nested-error structure. J. Amer. Statist. Assoc. 63, 626-632.

Greene, W. H. (1990). Econometric Analysis. McMillan Publishing Company, New York.

Harville, D. A. (1977). Maximum liklihood approachs to variance components estimation and to related problems. J. Amer. Statist. Assoc. 63, 320-336.

Henderson, C. R. (1953). Estimation of variance and covariance components. Biometrika 9, 226-252.

Henderson, C. R. (1973). Sire evaluation and genetic trends. In Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush 10-41. Amer. Soc. Animal Sci. -Amer. Dairy Sci. Assn. -Poultry Sci. Assn., Champaign, Ill.

Meyer, K. (1985). Maximum liklihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics 41, 153-165.

Miller, J.J. (1977). Asymptotic properties of maximum liklihood estimates in the mixed model of the analysis of variance. Ann. Statist. 5, 746-762.

Rao, C.R. (1973). Linear statistical inference and its applications. (Wiley, New York, 2nd ed.).

Remadi, S., and Y. Amemiya (1992). Limiting distribution of roots with differential rates of convergence and its applications. Preprint series No 92-5, Department of Statistics, Iowa State University.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. Statist. Science 6, 15-51.

Schott, J.R. and J.G. Saw (1984). A multivariate one-way classification model with random effects. J. Multivariate Anal. 15, 1-12.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). Variance components. Wiley, New York.

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. Biometrics 22, 527-550.

PAPER III.

ASYMPTOTIC PROPERTIES OF THE ESTIMATORS FOR MULTIVARIATE COMPONENTS OF VARIANCE

ABSTRACT

Estimation of the covariance matrices in the multivariate one-way random effect model is discussed. The rank of the between-group covariance matrix plays large role in model building as well as in assessing asymptotic properties of the estimated covariance matrices. The restricted (residual) maximum likelihood estimators derived under a rank condition are considered. Asymptotic properties of the estimators are derived for a possibly incorrectly specified rank and under either the number of groups, the number of replicates, or both tending to infinity. A higher order expansion covering various cases leads to a common approximate inference procedure which can be used in a wide range of practical situations. A simulation study supporting asymptotic theory is also presented.

1. INTRODUCTION

Suppose that a $p \times 1$ observation vector $\mathbf{Y}_{ij}$ taken on the $j-th$ individual in the $i-th$ group satisfies

$$\mathbf{Y}_{ij} = \mu + \mathbf{b}_i + \mathbf{w}_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, r, \tag{1.1}$$

where μ is a $p \times 1$ vector of unknown parameters, the $p \times 1$ $\mathbf{b}_i$ represents the $i-th$ group effect, and $\mathbf{w}_{ij}$ is the $p \times 1$ error term assumed to have $N_p(0,\ \Sigma_{ww})$ distribution. When n groups are assumed to be taken from some population, we often assume that the between-group effects $\mathbf{b}_i$'s are independent $N_p(\mathbf{0},\ \Sigma_{bb})$ random vectors distributed independently from the $\mathbf{w}_{ij}$'s. Assume that Σ_{bb} is nonnegative definite and Σ_{ww} is positive definite, and that $n > 1$ and $r > 1$. This is the multivariate one-way components of variance model.

The univariate components of variance model has been used and discussed extensively in the literature. For reviews, see, e.g., Harville (1977), Robinson (1991), and Searle et al. (1992). A multivariate model such as (1.1) containing unknown covariance matrices can be transformed to the general univariate form by stacking the $p \times 1$ response vectors. However, such re-writing may not solve some of the statistical problems for the multivariate case, because of the parameter space restriction and of the possibility of singular covariance matrices. Thus, development of statistical procedures for multivariate models often requires approaches slightly different from those used for the univariate case. The literature on the multivariate model has been rather limited. The ordinary or residual maximum likelihood estimation for model (1.1) with no restriction on Σ_{bb} was discussed by Klotz and Putter (1969), Bock and Petersen (1975), and Amemiya (1985). Hill and Thompson (1978) and Bhargava and Disch (1982) discuss the problem of a possible singular estimate of a covariance component. Estimation under the rank condition and testing for the rank was treated in Anderson (1984, 1989), Amemiya and Fuller (1984), Schott and Saw (1984), Anderson et al. (1986), Amemiya et al. (1990), and Anderson and Amemiya (1991). Thompson (1973) and Meyer (1985) discuss algorithms for computing the restricted (residual) maximum likelihood estimators (REML) of covariance components in the multivariate mixed effect model. For the balanced multivariate random effect model, Calvin and Dykstra (1991) proposed a computational algorithm which is guaranteed to converge to the REML. Properties of estimators or inference procedures for functions of covariance components have received virtually no treatment in the literature.

Here we consider estimation of covariance components Σ_{bb} and Σ_{ww} in model (1.1). Although model (1.1) is the simplest multivariate components of variance models, properties of the estimators have been largely unknown. Consideration of model (1.1) highlights some of the common problems for multivariate models, and suggests possible extensions to more general multivariate models. In model (1.1), the between-group effect $\mathbf{b}_i$ is $p \times 1$, i.e., each of the p response variables has one corresponding group effect variable in $\mathbf{b}_i$. But, the actual between-group variability can be concentrated in a space of dimension less than p. For example, some of the p variables or some linear combinations may have no between-group differences. Thus, a random effect in the multivariate model can exist with a singular covariance matrix with various values of rank, while a variance component in the univariate model is either zero or positive. As shown later, the true rank of a covariance component also affects properties of an estimated covariance component. In this paper, we consider for model (1.1), properties of estimators of Σ_{bb} and Σ_{ww} obtained under the assumption that rank $\Sigma_{bb} \leq m$. Note that any estimator should take values in (the closure of) the parameter space (with probability one), i.e., an estimator of Σ_{bb} should be a symmetric nonnegative definite matrix of rank at most m, and an estimator of Σ_{ww} should be symmetric positive definite, with probability one. A set of such estimators is the restricted (residual) maximum likelihood (REML) estimators derived under the assumption of rank $\Sigma_{bb} \leq m$. To present the estimators, let the between-group and within-group mean square matrices be defined to be

$$\begin{aligned} \mathbf{m}_{bb} &= \frac{r}{n-1}\sum_{i=1}^{n}(\bar{\mathbf{Y}}_{i.}-\bar{\mathbf{Y}}_{..})(\bar{\mathbf{Y}}_{i.}-\bar{\mathbf{Y}}_{..})', \\ \mathbf{m}_{ww} &= \frac{1}{n(r-1)}\sum_{i=1}^{n}\sum_{j=1}^{r}(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_{i.})(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_{i.})', \end{aligned} \tag{1.2}$$

where $\mathbf{Y}_{i.} = \frac{1}{r}\sum_{j=1}^{r}\mathbf{Y}_{ij}$ and $\mathbf{Y}_{..} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}\mathbf{Y}_{ij}$. Note that the statistic $\frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ is unbiased for Σ_{bb}, but does not always take values in the parameter space. Let a $p \times p$ orthogonal $\hat{\mathbf{Q}}$ and $p \times p$ diagonal $\hat{\Lambda} =$ diag $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_p\}$ be such that

$$\begin{aligned} \mathbf{m}_{ww}^{-1/2}\mathbf{m}_{bb}\mathbf{m}_{ww}^{-1/2} &= \hat{\mathbf{Q}}\hat{\Lambda}\hat{\mathbf{Q}}', \\ \hat{\lambda}_1 \geq \ldots &\geq \hat{\lambda}_p. \end{aligned} \tag{1.3}$$

Define $\hat{k} =$ min $\{m,$ number of $\hat{\lambda}_i$'s $> 1\}$. This $\hat{k}$ will be the rank of the REML estimator $\hat{\Sigma}_{bb}$. We write

$$\hat{\mathbf{P}} = (\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2) = \mathbf{m}_{ww}^{1/2}\hat{\mathbf{Q}},$$

$$\hat{\mathbf{\Lambda}} = \begin{pmatrix} \hat{\mathbf{\Lambda}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Lambda}}_2 \end{pmatrix}, \tag{1.4}$$

where $\hat{\mathbf{P}}_1$ is $p \times \hat{k}$, and the $\hat{k} \times \hat{k}$ $\hat{\mathbf{\Lambda}}_1$ consists of the $\hat{k}$ largest roots $\hat{\lambda}_i$'s. Then, the REML estimators under rank $\mathbf{\Sigma}_{bb} < m$ are

$$\begin{aligned} \hat{\mathbf{\Sigma}}_{bb}(m) &= \mathbf{0}, && \text{if } \hat{k} = 0, \\ &= \frac{1}{r}\hat{\mathbf{P}}_1(\hat{\mathbf{\Lambda}}_1 - \mathbf{I}_{\hat{k}})\hat{\mathbf{P}}_1', && \text{if } \hat{k} > 0, \\ \hat{\mathbf{\Sigma}}_{ww}(m) &= \frac{1}{nr-1}\{(n-1)[\mathbf{m}_{bb} - r\hat{\mathbf{\Sigma}}_{bb}] + n(r-1)\mathbf{m}_{ww}\}. \end{aligned} \tag{1.5}$$

See, e.g., Anderson (1984) and Amemiya and Fuller (1984). Note that $\hat{\mathbf{\Sigma}}_{bb}(m)$ is a symmetric nonnegative definite matrix of rank $\hat{k} \leq m$ and that $\hat{\mathbf{\Sigma}}_{ww}(m)$ is a weighted average of $\mathbf{m}_{ww}$ and a part of $\mathbf{m}_{bb}$ not used for estimating $\mathbf{\Sigma}_{bb}$. An alternative form of the REML estimators is

$$\begin{aligned} \hat{\mathbf{\Sigma}}_{bb}(m) &= \frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww}) - \frac{1}{r}\hat{\mathbf{\Omega}}_m, \\ \hat{\mathbf{\Sigma}}_{ww}(m) &= \mathbf{m}_{ww} + \frac{n-1}{nr-1}\hat{\mathbf{\Omega}}_m, \end{aligned} \tag{1.6}$$

where

$$\hat{\mathbf{\Omega}}_m = \hat{\mathbf{P}}_2(\hat{\mathbf{\Lambda}}_2 - \mathbf{I}_{p-\hat{k}})\hat{\mathbf{P}}_2'.$$

In this form, we see that $\hat{\mathbf{\Sigma}}_{bb}(m)$ and $\hat{\mathbf{\Sigma}}_{ww}(m)$ are obtained by adjusting the unbiased statistic $\frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ and $\mathbf{m}_{ww}$ using terms involving $\hat{\mathbf{\Omega}}_m$, and that a partition of the total sum squares holds;

$$\begin{aligned} (n-1)r\hat{\mathbf{\Sigma}}_{bb}(m) + n(r-1)\hat{\mathbf{\Sigma}}_{ww}(m) &= (n-1)\mathbf{m}_{bb} + n(r-1)\mathbf{m}_{ww} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{r}(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_{..})'. \end{aligned} \tag{1.7}$$

In this paper, we derive asymptotic properties of $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$. Development of asymptotic results for the random effects models require some special care. See, e.g., Miller (1977). Model (1.1) contains two indices n and r, i.e., the numbers of groups and replicates. A practical situation may have large n, large r, or large n and r. We develop asymptotic theory covering any of these cases. Throughout this paper, we refer to our different assumptions for asymptotics as

case I: $n \to \infty$ and r is fixed,

case II: $n \to \infty$, and $r \to \infty$,

case III: n is fixed and $r \to \infty$.

Our eventual goal is to develop approximate inference procedures for Σ_{bb} and Σ_{ww} or (functions of Σ_{bb} and Σ_{ww})which can be applied in a wide range of practical situations. After developing specific results for each of these cases, we will suggest approximate inference procedure which can be used in a situation corresponding to any one of cases I, II, and III. Another problem associated with developing asymptotic properties of $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$ is their dependency on the true rank m_0 of Σ_{bb}. Although the estimators are obtained under the assumption that rank $\Sigma_{bb} \leq m$, the true rank m_0 is generally unknown. We will investigate the effect of not knowing m_0 on the properties of the estimators.

2. CONSISTENCY

To discuss asymptotic properties of $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$, we need to note that for case III with fixed n, a consistent estimator of Σ_{bb} does not exist. Thus, for $\hat{\Sigma}_{bb}(m)$, we discuss the consistency by checking whether or not

$$\hat{\Sigma}_{bb}(m) - \mathbf{S}_{bb} \xrightarrow{P} \mathbf{0}, \tag{2.1}$$

where

$$\begin{aligned} \mathbf{S}_{bb} &= \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{b}_i - \bar{\mathbf{b}})(\mathbf{b}_i - \bar{\mathbf{b}})', \\ \bar{\mathbf{b}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{b}_i. \end{aligned}$$

Note that this is equivalent to the ordinary consistency for cases I and II with $n \to \infty$. First, we consider the case with $m \geq m_0$, i.e., where the maximum allowable rank is larger or equal to the true rank of Σ_{bb}.

Theorem 1. If $m \geq m_0$, then for all cases I, II, and III

$$\begin{aligned} \hat{\Sigma}_{bb}(m) &= \frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww}) + O_P(\frac{1}{r\sqrt{n}}) \\ &= \mathbf{S}_{bb} + O_P(\frac{1}{\sqrt{rn}}), \\ \hat{\Sigma}_{ww}(m) &= \mathbf{m}_{ww} + O_P(\frac{1}{r\sqrt{n}}) \\ &= \Sigma_{ww} + O_P(\frac{1}{\sqrt{rn}}), \end{aligned}$$

where $\mathbf{S}_{bb}$ is defined in (2.1), r is constant for case I, and n is constant for case III.

Proof. Note that for all three cases,

$$
\begin{aligned}
\frac{1}{r}\mathbf{m}_{bb} - \mathbf{S}_{bb} - \frac{1}{r}\Sigma_{ww} &= O_p(\frac{1}{\sqrt{nr}}), \\
\mathbf{m}_{ww} - \Sigma_{ww} &= O_p(\frac{1}{\sqrt{nr}}), \\
\hat{\mathbf{P}} = m_{ww}^{1/2}\hat{\mathbf{Q}} &= \Sigma_{ww}^{1/2}\hat{\mathbf{Q}} + O_p(\frac{1}{\sqrt{nr}}) \\
&= O_p(1),
\end{aligned}
\tag{2.2}
$$

where we have used the fact that the elements of $\hat{\mathbf{Q}}$ are bounded by one in absolute value. By the result on the limiting distribution of the roots $\hat{\lambda}_i$'s.

$$\sqrt{n}(\hat{\lambda}_i - 1) = O_p(1), \; i = m_0 + 1, \ldots, p. \tag{2.3}$$

Note that $\hat{\Omega}_m$ in (1.6) is a function of $\hat{\lambda}_i$, $i = m+1, \ldots, p$ with $m \geq m_0$. It follows from (2.2) and (2.3) that for all three cases

$$\hat{\Omega}_m = O_p(\frac{1}{\sqrt{n}}). \tag{2.4}$$

Thus the result follows from (1.6), (2.2), and (2.4).

Hence, with the understanding of the consistency of $\hat{\Sigma}_{bb}(m)$ as given in (2.1) for case III, $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$ are consistent for Σ_{bb} and Σ_{ww}, provided that $m \geq m_0$. Thus, for example, $\Sigma_{bb}(p)$ and $\Sigma_{ww}(p)$ obtained under no rank condition of Σ_{bb} are always consistent.

To discuss the case $m < m_0$, define the $p \times p$ matrix

$$
\Psi = \begin{cases}
\Sigma_{ww}^{-1/2}\Sigma_{bb}\Sigma_{ww}^{-1/2}, & \text{for cases I and II,} \\
\Sigma_{ww}^{-1/2}\mathbf{S}_{bb}\Sigma_{ww}^{-1/2}, & \text{for case III.}
\end{cases}
\tag{2.5}
$$

Note that $\mathbf{S}_{bb}$ is a random matrix defined in (2.1). Let $\nu_1 \geq \ldots \geq \nu_p$ be the eigenvalues of $\mathbf{\Psi}$, and let $\mathbf{Q}_0$ be the $p \times (m_0 - m)$ matrix of a set of eigenvectors corresponding to $\nu_{m+1}, \ldots, \nu_{m_0}$. For case III, ν_i's and $\mathbf{Q}_0$ are also random. By Okamoto (1973), for case III, ν_i's are distinct with probability one. For simplicity, we assume for cases I and II that $\nu_m > \nu_{m+1}$. Now we present the following result on the consistency when $m < m_0$.

Theorem 2. If $m < m_0$, then for all cases I, II, and III,

$$\hat{\mathbf{\Sigma}}_{bb}(m) - \mathbf{S}_{bb} \xrightarrow{P} -\mathbf{B}$$

$$\hat{\Sigma}_{ww}(m) - \Sigma_{ww} \xrightarrow{P} \begin{cases} \mathbf{B} & \text{, for cases I and II,} \\ \frac{n-1}{n}\mathbf{B} & \text{, for case III,} \end{cases}$$

where

$$\mathbf{B} = \mathbf{Q}_0 \text{ diag } \{\nu_{m+1}, \ldots, \nu_{m_0}\}\mathbf{Q}_0'.$$

Proof. It follows from (2.2) that for all cases I, II, and III

$$\frac{1}{r}\mathbf{m}_{ww}^{-1/2}\mathbf{m}_{bb}\mathbf{m}_{ww}^{-1/2} \xrightarrow{a.s.} \mathbf{\Psi}_0, \tag{2.6}$$

where

$$\mathbf{\Psi}_0 = \begin{cases} \mathbf{\Psi} + \frac{1}{r}\mathbf{I}_p & \text{, for case I,} \\ \mathbf{\Psi} & \text{, for cases II and III.} \end{cases}$$

Thus, by the continuity of the eigenvalues

$$\frac{1}{r}\hat{\lambda}_i \xrightarrow{a.s} \nu_i^0, \ i = 1, 2, \ldots, p, \tag{2.7}$$

where $\nu_i^0 = \nu_i + \frac{1}{r}$ for case I and $\nu_i^0 = \nu_i$ for cases II and III. Let $\hat{\mathbf{Q}}_0$ be the $(m+1) - st$ through $m_0 - th$ columns of $\hat{\mathbf{Q}}$. Since the elements of $\hat{\mathbf{Q}}_0$ are bounded, with probability one, every subsequence has a converging subsubsequence. By (2.6) and (2.7) over such a converging subsubsequence $\hat{\mathbf{Q}}_0$ with a limit $\mathbf{Q}_0^*$

$$\begin{aligned}\mathbf{0} &= \frac{1}{r}\mathbf{m}_{ww}^{-1/2}\mathbf{m}_{bb}\mathbf{m}_{ww}^{-1/2}\hat{\mathbf{Q}}_m - \hat{\mathbf{Q}}_m\frac{1}{r}\text{ diag }\{\hat{\lambda}_{m+1},\ldots,\hat{\lambda}_{m_0}\}\\ &\to \boldsymbol{\Psi}_0\mathbf{Q}_0^* - \mathbf{Q}_0^*\text{ diag }\{\nu_{m+1}^0,\ldots,\nu_{m_0}^0\}.\end{aligned}$$

Because $\nu_m^0 > \nu_{m+1}^0$ and $\nu_{m_0}^0 > \nu_{m_0+1}^0$ for cases I and II, and for case III with probability one, $\mathbf{Q}_0^*$ is unique up to orthogonal rotation of each eigenspace of $\boldsymbol{\Psi}_0$ corresponding to $\nu_{m+1}^0,\ldots,\nu_{m_0}^0$. Since $\mathbf{Q}_0^*\mathbf{Q}_0^{*'}$ and $\mathbf{Q}_0^*\text{ diag}\{\nu_{m+1}^0,\ldots,\nu_{m_0}^0\}\mathbf{Q}_0^{*'}$ are unique under such orthogonal rotations, and equal to $\mathbf{Q}_0\mathbf{Q}_0'$ and $\mathbf{Q}_0\text{ diag}\{\nu_{m+1}^0,\ldots,\nu_{m_0}^0\}\mathbf{Q}_0'$, it follows that for all three cases

$$\mathbf{R}_m = \hat{\mathbf{Q}}_0\frac{1}{r}\text{ diag}\{\hat{\lambda}_{m+1}-1,\ldots,\hat{\lambda}_{m_0}-1\}\hat{\mathbf{Q}}_0' \xrightarrow{a.s} \mathbf{B} \tag{2.8}$$

Since $\nu_{m_0}^0 > 1$ for case I, and since $r \to \infty$ for cases II and III, (2.7) implies that for all three cases

$$P\{\hat{\lambda}_{m_0} > 1\} \to 1. \tag{2.9}$$

For $m < m_0$

$$P\{\hat{k} = m\} \to 1. \tag{2.10}$$

Hence, using the form (1.6), we can write with probability approaching one,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{bb}(m) &= \hat{\boldsymbol{\Sigma}}_{bb}^0(m_0) - \frac{1}{r}\mathbf{R}_m,\\ \hat{\boldsymbol{\Sigma}}_{ww}(m) &= \hat{\boldsymbol{\Sigma}}_{ww}^0(m_0) - \frac{n-1}{nr-1}\mathbf{R}_m,\end{aligned} \tag{2.11}$$

where $\hat{\boldsymbol{\Sigma}}_{bb}^0(m_0)$ and $\hat{\boldsymbol{\Sigma}}_{ww}^0(m_0)$ are $\hat{\boldsymbol{\Sigma}}_{bb}(m_0)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m_0)$ with $\hat{k} = m_0$, and $\mathbf{R}_m$ is defined in (2.8). Thus, the result follows from (2.8), (2.11) and Theorem 1.

Hence, $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$ are not consistent when the specified maximum rank m for Σ_{bb} is smaller than the true rank m_0. Since the matrix $\mathbf{B}$ in Theorem 2 is nonnegative definite, $\hat{\Sigma}_{bb}(m)$ "underestimates" Σ_{bb} and $\hat{\Sigma}_{ww}(m)$ "overestimates" Σ_{ww}. Thus, it is important not to underspecify the rank of Σ_{bb} in estimation of Σ_{bb} and Σ_{ww}.

3. LIMITING DISTRIBUTION

By Theorem 2, if $m < m_0$, $\hat{\Sigma}_{bb}(m)$ and $\hat{\Sigma}_{ww}(m)$ are inconsistent in the sense given in the theorem. For such a case, a limiting distribution result useful for asymptotic inferences can not easily be found. Hence, we consider only the case with $m \geq m_0$, i.e., where the true rank is less than or equal to the assumed maximum rank. For a $p \times p$ symmetric matrix $\mathbf{A}$, we use the notation $vech\mathbf{A}$, a $p(p+1)/2 \times 1$ vector containing the elements on and below the diagonal of $\mathbf{A}$ starting with the first column. For a $p \times p$ symmetric matrix $\mathbf{A}$, there is a unique $p^2 \times p(p+1)/2$ matrix $\mathbf{K}_p$ such that $vec\mathbf{A} = \mathbf{K}_p vech\mathbf{A}$, where $vec\mathbf{A}$ is the $p^2 \times 1$ vector listing the elements of the columns of $\mathbf{A}$ starting with the first. For any such $\mathbf{A}$ we write

$$\mathbf{\Gamma}(A) = 2\mathbf{K}_p^{+}(\mathbf{A} \bigotimes \mathbf{A})\mathbf{K}_p^{+'}, \tag{3.1}$$

where $\mathbf{K}_p^{+} = (\mathbf{K}_p'\mathbf{K}_p)^{-1}\mathbf{K}_p$ and $\otimes$ is the Kronecker product. Note that for $\mathbf{A} = (a_{ij})$, a typical element of $\mathbf{\Gamma}(\mathbf{A})$ is $a_{ik}a_{jl} + a_{il}a_{jk}$.

If $m \geq m_0$ the limiting distributions for cases II and III are relatively simple and are given in the following Theorem.

Theorem 2. Suppose that $m \geq m_0$. For case II

$$\left\{ \begin{array}{c} \sqrt{n} vech(\hat{\mathbf{\Sigma}}_{bb}(m) - \mathbf{\Sigma}_{bb}) \\ \sqrt{n(r-1)} vech(\hat{\mathbf{\Sigma}}_{ww}(m) - \mathbf{\Sigma}_{ww}) \end{array} \right\} \xrightarrow{L} N \left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{\Gamma}(\mathbf{\Sigma}_{bb}) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}(\mathbf{\Sigma}_{ww}) \end{pmatrix} \right\}.$$

For case III

$$\left\{ \begin{array}{c} vech\hat{\boldsymbol{\Sigma}}_{bb}(m) \\ \sqrt{n(r-1)}vech(\hat{\boldsymbol{\Sigma}}_{ww}(m) - \boldsymbol{\Sigma}_{ww}) \end{array} \right\} \xrightarrow{L} \left\{ \begin{array}{c} vech\mathbf{S}_{bb} \\ \mathbf{W} \end{array} \right\},$$

where $\mathbf{S}_{bb}$ and $\mathbf{W}$ are independent, $(r-1)\mathbf{S}_{bb} \sim W_p(\boldsymbol{\Sigma}_{bb},\ n-1)$, and $\mathbf{W} \sim N(\mathbf{0},\ \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{ww}))$.

Proof. The results follow immediately from Theorem 1.

Note that the limiting distribution of $\hat{\boldsymbol{\Sigma}}_{bb}(m)$ and $\hat{\boldsymbol{\Sigma}}_{ww}(m)$ in Theorem 3 is that of $\mathbf{S}_{bb}$ and $\mathbf{m}_{ww}$. Thus, for cases II and III with $r \to \infty$, the limiting distributions are simple in that the rather complex nature of the rank restriction and inter-relationship disappear in the limit. Also, the form of these limiting distributions are the same for all $m \geq m_0$.

For case I, we need to distinguish two cases, $m = m_0$ and $m > m_0$. For rank $\boldsymbol{\Sigma}_{bb} = m_0 < p$, let $\mathbf{C}$ be the $p \times (p - m_0)$ matrix of rank $(p - m_0)$ such that $\mathbf{C}'\boldsymbol{\Sigma}_{bb}\mathbf{C} = \mathbf{0}$. Define

$$\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_{ww}\mathbf{C}(\mathbf{C}'\boldsymbol{\Sigma}_{ww}\mathbf{C})^{-1}\mathbf{C}'\boldsymbol{\Sigma}_{ww}. \tag{3.2}$$

Note that $\mathbf{C}$ is not unique but $\boldsymbol{\Sigma}_0$ is free of the choice of $\mathbf{C}$. If $m_0 = p$, $\boldsymbol{\Sigma}_0$ is understood to be zero.

Theorem 4. Consider case I. If $m = m_0$, then

$$\left\{ \begin{array}{c} \sqrt{n}vech(\hat{\boldsymbol{\Sigma}}_{bb}(m_0) - \boldsymbol{\Sigma}_{bb}) \\ \sqrt{n(r-1)}vech(\hat{\boldsymbol{\Sigma}}_{ww}(m_0) - \boldsymbol{\Sigma}_{ww}) \end{array} \right\} \xrightarrow{L} N\left\{ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}^{\mathrm{I}}_{bb} & \mathbf{V}^{\mathrm{I}}_{bw} \\ \mathbf{V}^{\mathrm{I}}_{wb} & \mathbf{V}^{\mathrm{I}}_{ww} \end{pmatrix} \right\},$$

where

$$\mathbf{V}^{\mathrm{I}}_{bb} = \boldsymbol{\Gamma}(\boldsymbol{\Sigma}_{bb} + \frac{1}{r}\boldsymbol{\Sigma}_{ww}) + \frac{1}{r-1}\boldsymbol{\Gamma}(\frac{1}{r}\boldsymbol{\Sigma}_{ww}) - \frac{1}{r(r-1)}\boldsymbol{\Gamma}(\boldsymbol{\Sigma}_0),$$

$$\begin{aligned}\mathbf{V}^{\mathrm{I}}_{bw} &= \frac{1}{r\sqrt{r-1}}[\mathbf{\Gamma}(\mathbf{\Sigma}_0) - \mathbf{\Gamma}(\mathbf{\Sigma}_{ww})], \\ \mathbf{V}^{\mathrm{I}}_{ww} &= \mathbf{\Gamma}(\mathbf{\Sigma}_{ww}) - \frac{1}{r}\mathbf{\Gamma}(\mathbf{\Sigma}_0).\end{aligned}$$

If $m > m_0$, the limiting distribution of

$$\left\{\begin{array}{c}\sqrt{n}vech(\hat{\mathbf{\Sigma}}_{bb}(m) - \mathbf{\Sigma}_{bb}) \\ \sqrt{n(r-1)}vech(\hat{\mathbf{\Sigma}}_{ww}(m) - \mathbf{\Sigma}_{ww})\end{array}\right\}$$

either does not exist or is not normal.

Proof. The result for $m = m_0$ is a consequence of the expansion given in Theorem 6 of the next section. If $m > m_0$, for $\hat{\mathbf{\Sigma}}_{bb}$, for example,

$$\sqrt{n}(\hat{\mathbf{\Sigma}}_{bb}(m) - \mathbf{\Sigma}_{bb}) = \sqrt{n}(\hat{\mathbf{\Sigma}}^0_{bb}(m_0) - \mathbf{\Sigma}_{bb}) + \frac{1}{r}\mathbf{m}^{1/2}_{ww}\sqrt{n}\mathbf{R}\mathbf{m}^{1/2}_{ww}, \tag{3.3}$$

where $\hat{\mathbf{\Sigma}}^0_{bb}(m_0)$ is $\hat{\mathbf{\Sigma}}_{bb}(m_0)$ with $\hat{k} = m_0$,

$$\mathbf{R} = \sum_{i=m_0+1}^{m} (\hat{\lambda}_i - 1)I(\hat{\lambda}_i > 1)\hat{\mathbf{q}}_i\hat{\mathbf{q}}_i',$$

and $\hat{\mathbf{q}}_i$ is the $i-th$ column of $\hat{\mathbf{Q}}$. Since $\hat{\mathbf{\Sigma}}^0_{bb}(m_0)$ is a function of $\hat{\lambda}_i$, $i = 1, 2 \ldots, m_0$, and since $\mathbf{m}_{ww} \xrightarrow{P} \mathbf{\Sigma}_{ww}$, the two terms in (3.3) are independent in the limit. By the first part of this theorem, the first term converges to a normal distribution. Thus for the sum to have a limiting normal distribution, the second term must have a limiting normal distribution. See, e.g., Feller(1971, p.525, Cramer-Levy theorem). Note that

$$\sqrt{n}(\hat{\lambda}_i - 1)I(\sqrt{n}(\hat{\lambda}_i - 1) > 0),\ i = m_0 + 1, \ldots, m,$$

converge to nonnegative random variables. Thus, if $\sqrt{n}\mathbf{R}$ has a nondegenerate limiting distribution, then it is not normal. For $m < p$, following the argument used in the proof of Theorem 2, the limit of a converging subsequence of $\sqrt{n}\mathbf{R}$ depends on the subsequence, because the limiting $\hat{q}_i$'s span the eigenspace of dimension $(p - m_0)$ corresponding to the unit root of $\Sigma_{ww}^{-1/2}(r\Sigma_{bb} + \Sigma_{ww})\Sigma_{ww}^{-1/2}$. Hence, for $m < p$, $\sqrt{n}(\hat{\Sigma}_{bb}(m) - \Sigma_{bb})$ does not have a limiting normal distribution.

Thus, for case I, the limiting distribution for $m = m_0$ is not a valid limiting distribution for $m > m_0$. For $m > m_0$, the estimators are consistent, but do not have a limiting normal distribution.

The discrepancies among the limiting distributions for cases I, II, and III as given in Theorems 3 and 4 show that the use of the limiting result for asymptotic inferences requires some special care in practice. One needs to decide which of the three cases is most appropriate for a given situation.

4. ASYMPTOTIC APPROXIMATION

A possible approach to developing approximate inference procedures useful for all cases I, II, and III is to consider a common asymptotic expansion. To this end, we assume that $m = m_0$, and first derive an expansion for each of the three cases which is of higher order than that given in Theorem 1. We recall the definition of the $p \times (p - m_0)$ matrix $\mathbf{C}$ in (3.2). Let $\mathbf{D}$ be a $p \times m_0$ matrix of rank m_0 satisfying that $(\mathbf{C}, \mathbf{D})$ is a $p \times p$ nonsingular matrix. Note that $\mathbf{D}$ is unique only up to multiplication of a $m_0 \times m_0$ nonsingular matrix from the right, and that $\mathbf{D}'\Sigma_{bb}\mathbf{D}$ is nonsingular. Define

$$\begin{aligned}
\mathbf{S}_{xx} &= \mathbf{D}'\mathbf{S}_{bb}\mathbf{D}, \\
\mathbf{S}_{vv} &= \mathbf{C}'\frac{r}{n-1}\sum_{i=1}^{n}(\bar{\mathbf{w}}_{i.} - \bar{\mathbf{w}}_{..})(\bar{\mathbf{w}}_{i.} - \bar{\mathbf{w}}_{..})'\mathbf{C}, \\
\mathbf{S}_{vx} &= \mathbf{C}'\frac{\sqrt{r}}{n-1}\sum_{i=1}^{n}(\bar{\mathbf{w}}_{i.} - \bar{\mathbf{w}}_{..})(\mathbf{b}_i - \bar{\mathbf{b}}_{.})'\mathbf{D}', \\
\mathbf{S}_{ee} &= \mathbf{C}'\frac{1}{n(r-1)}\sum_{i=1}^{n}\sum_{j=1}^{r}(\mathbf{w}_{ij} - \bar{\mathbf{w}}_{i.})(\mathbf{w}_{ij} - \bar{\mathbf{w}}_{i.})'\mathbf{C},
\end{aligned} \tag{4.1}$$

where $\bar{\mathbf{w}}_{i.} = \frac{1}{r}\sum_{i=1}^{r}\mathbf{w}_{ij}$, $\bar{\mathbf{w}}_{..} = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{w}}_{i.}$ and $\bar{\mathbf{b}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{b}_i$. Let

$$\begin{aligned}
\Omega_{\mathrm{I}} &= \Sigma_{ww}\mathbf{C}(\mathbf{C}'\Sigma_{ww}\mathbf{C})^{-1}(\mathbf{S}_{vv} - \mathbf{S}_{ee})(\mathbf{C}'\Sigma_{ww}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{ww}, \\
\Omega_{\mathrm{II}} &= \Sigma_{ww}\mathbf{C}(\mathbf{C}'\Sigma_{ww}\mathbf{C})^{-1}(\mathbf{S}_{vv.x} - \mathbf{S}_{ee})(\mathbf{C}'\Sigma_{ww}\mathbf{C})^{-1}\mathbf{C}'\Sigma_{ww},
\end{aligned} \tag{4.2}$$

where

$$\mathbf{S}_{vv.x} = \mathbf{S}_{vv} - \mathbf{S}_{vx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{vx}'.$$

Note that $\mathbf{\Omega}_{\mathrm{I}}$ and $\mathbf{\Omega}_{\mathrm{II}}$ are invariant for different choices for $\mathbf{C}$ and $\mathbf{D}$.

Theorem 5. Suppose that $m = m_0 < p$. Then, for case I,

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_{bb}(m_0) - \mathbf{\Sigma}_{bb} &= \frac{1}{r}[\mathbf{m}_{bb} - (r\mathbf{\Sigma}_{bb} + \mathbf{\Sigma}_{ww})] - \frac{1}{r}(\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) - \frac{1}{r}\mathbf{\Omega}_{\mathrm{I}} + O_p(\frac{1}{n}), \\
\hat{\mathbf{\Sigma}}_{ww}(m_0) - \mathbf{\Sigma}_{ww} &= (\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) + \frac{n-1}{nr-1}\mathbf{\Omega}_{\mathrm{I}} + O_p(\frac{1}{n}).
\end{aligned}$$

For case II,

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_{bb}(m_0) - \mathbf{\Sigma}_{bb} &= \frac{1}{r}[\mathbf{m}_{bb} - (r\mathbf{\Sigma}_{bb} + \mathbf{\Sigma}_{ww})] - \frac{1}{r}(\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) - \frac{1}{r}\mathbf{\Omega}_{\mathrm{II}} \\
&\quad + O_p(\frac{1}{nr\sqrt{r}}), \\
\hat{\mathbf{\Sigma}}_{ww}(m_0) - \mathbf{\Sigma}_{ww} &= (\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) + \frac{n-1}{nr-1}\mathbf{\Omega}_{\mathrm{II}} + O_p(\frac{1}{nr\sqrt{r}}).
\end{aligned}$$

For case III,

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_{bb}(m_0) - \mathbf{\Sigma}_{bb} &= \frac{1}{r}[\mathbf{m}_{bb} - (r\mathbf{\Sigma}_{bb} + \mathbf{\Sigma}_{ww})] - \frac{1}{r}\mathbf{\Omega}_{\mathrm{II}} + O_p(\frac{1}{r\sqrt{r}}), \\
\hat{\mathbf{\Sigma}}_{ww}(m_0) - \mathbf{\Sigma}_{ww} &= (\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) + \frac{n-1}{nr-1}\mathbf{\Omega}_{\mathrm{II}} + O_p(\frac{1}{r\sqrt{r}}).
\end{aligned}$$

The proof of Theorem 5 is given in section 7. These three expansions are meaningful in the sense that every term explicitly given is of order larger than the remainder. All explicit terms for case I are $O_p(\frac{1}{n})$. But, for cases II and III, terms have different order, representing higher order expansions. Since the expansions for the three cases are similar, an expansion valid for all cases can be derived.

Theorem 6. If $m = m_0$, then for all cases I, II, and III,

$$\begin{aligned}\hat{\Sigma}_{bb}(m_0) - \Sigma_{bb} &= \frac{1}{r}[\mathbf{m}_{bb} - (r\Sigma_{bb} + \Sigma_{ww})] - \frac{1}{r}(\mathbf{m}_{ww} - \Sigma_{ww}) - \frac{1}{r}\Omega_{II} \\ &\quad + O_p(\frac{1}{nr\sqrt{r}}), \\ \hat{\Sigma}_{ww}(m_0) - \Sigma_{ww} &= (\mathbf{m}_{ww} - \Sigma_{ww}) + \frac{n-1}{nr-1}\Omega_{II} + O_p(\frac{1}{nr\sqrt{r}}),\end{aligned}$$

where it is understood that n is constant for case III, r is constant for case I, and that $\Omega_{II} = \mathbf{0}$ for $m_0 = p$.

Proof. Note that for case I

$$\mathbf{S}_{vv} - \mathbf{S}_{vv.x} = \mathbf{S}_{vx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{vx}' = O_p(\frac{1}{n}),$$

and thus $\Omega_I - \Omega_{II} = O_p(\frac{1}{n})$. For case III

$$\frac{1}{r}(\mathbf{m}_{ww} - \Sigma_{ww}) = O_p(\frac{1}{r\sqrt{r}}).$$

Thus, the result follows.

The common expansion given in Theorem 6 is in fact the one for case II given in Theorem 5. For cases I and III, this common expansion simply adds extra terms of the same order as the remainder. This expansion also highlights some characteristics of the estimators derived under the rank condition much better than the expansion in Theorem 1. The adjustment or improvement made to the naive estimators $\frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ and $\mathbf{m}_{ww}$ are given in terms of Ω_{II}. The term Ω_{II} can be characterized to be a part of $(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ estimating the error variability, not the between-group variability. This term is subtracted from $\frac{1}{r}(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ for an efficient estimator

$\hat{\Sigma}_{bb}(m)$, and is pooled with $\mathbf{m}_{ww}$ for an improved estimator of Σ_{ww}. This expansion provides a means for obtaining an approximate inference procedure which works for a wide range of practical situations.

5. APPROXIMATE INFERENCE PROCEDURES

We develop approximate inference procedures for functions of the elements of Σ_{bb} or functions of the elements of Σ_{bb} and Σ_{ww}. Typical examples are a linear combination of the elements of Σ_{bb}, a between-group correlation (a correlation computed from Σ_{bb}), and an intra-group correlation (a diagonal element of Σ_{bb} divided by the sum of diagonal elements of Σ_{bb} and Σ_{ww}). In practice, it may be difficult to decide which of the three cases I, II, and III is most appropriate for a particular situation. Thus, our goal here is to develop procedures useful for various situations. From the results in sections 2 and 3, we note that without some knowledge of the rank of Σ_{bb}, inference procedures can be incorrect, especially for case I. Thus, if the rank is unknown we suggest performing some statistical inference for the rank. See Anderson (1989), Amemiya et al. (1990), Anderson and Amemiya (1991). Here, we assume that some idea about m_0, the true rank of Σ_{bb} is obtained so that $\hat{\Sigma}_{bb}(m_0)$ and $\hat{\Sigma}_{ww}(m_0)$ can be used at least with large enough probability. First we consider the covariance matrix of the terms in the common expansion in Theorem 6 as a common approximate covariance matrix. Taking the covariance matrix of the expansion terms and ignoring the remainder, the approximate covariance matrix of

$$\left\{ \begin{array}{l} vech\hat{\Sigma}_{bb}(m_0) \\ vech\hat{\Sigma}_{ww}(m_0) \end{array} \right\}$$

is

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{bb} & \mathbf{V}_{bw} \\ \mathbf{V}_{wb} & \mathbf{V}_{ww} \end{pmatrix},$$

where

$$\begin{aligned}
\mathbf{V}_{bb} &= \frac{1}{n-1}\mathbf{\Gamma}(\mathbf{\Sigma}_{bb} + \frac{1}{r}\mathbf{\Sigma}_{ww}) + \frac{1}{n(r-1)}\mathbf{\Gamma}(\frac{1}{r}\mathbf{\Sigma}_{ww}) \\
&\quad -\frac{1}{r^2}(\frac{1}{n-1} + \frac{1}{n(r-1)} + \frac{m_0}{(n-1)^2})\mathbf{\Gamma}(\mathbf{\Sigma}_0), \\
\mathbf{V}_{bw} &= \frac{1}{r}(\frac{1}{n(r-1)} + \frac{m_0}{(n-1)(nr-1)})\mathbf{\Gamma}(\mathbf{\Sigma}_0) - \frac{1}{rn(r-1)}\mathbf{\Gamma}(\mathbf{\Sigma}_{ww}), \\
\mathbf{V}_{ww} &= \frac{1}{n(r-1)}\mathbf{\Gamma}(\mathbf{\Sigma}_{ww}) - (\frac{n-1}{n(r-1)(nr-1)} + \frac{m_0}{(nr-1)^2})\mathbf{\Gamma}(\mathbf{\Sigma}_0),
\end{aligned}$$

the $\mathbf{\Gamma}(.)$ function and $\mathbf{\Sigma}_0$ are defined in (3.1) and (3.2), respectively. The unknown matrices $\mathbf{\Sigma}_{bb}$ and $\mathbf{\Sigma}_{ww}$ in $\mathbf{V}$ can be estimated by $\hat{\mathbf{\Sigma}}_{bb}(m_0)$ and $\hat{\mathbf{\Sigma}}_{ww}(m_0)$. To estimate $\mathbf{\Sigma}_0$, we consider any $p \times (p - \hat{k})$ $\hat{\mathbf{C}}$ such that $\hat{\mathbf{C}}'\hat{\mathbf{\Sigma}}_{bb}(m_0)\hat{\mathbf{C}} = 0$. Recall that rank $\hat{\mathbf{\Sigma}}_{bb}(m_0) = \hat{k}$. By (7.1) in the appendix, $P\{\hat{k} = m_0\} \to 1$ for all three cases. Thus, for all three cases,

$$\begin{aligned}
\hat{\mathbf{\Sigma}}_0 &= \hat{\mathbf{\Sigma}}_{ww}(m_0)\hat{\mathbf{C}}(\hat{\mathbf{C}}'\hat{\mathbf{\Sigma}}_{ww}(m_0)\hat{\mathbf{C}})^{-1}\hat{\mathbf{C}}'\hat{\mathbf{\Sigma}}_{ww}(m_0) \\
&\xrightarrow{P} \mathbf{\Sigma}_0.
\end{aligned}$$

Hence, we have an estimated covarinace matrix $\hat{\mathbf{V}}$ obtained by evaluating $\mathbf{V}$ at $\hat{\mathbf{\Sigma}}_{bb}(m_0)$, $\hat{\mathbf{\Sigma}}_{ww}(m_0)$, and $\hat{\mathbf{\Sigma}}_0$. We suggest the use of $\hat{\mathbf{\Sigma}}_{bb}(m)$, $\hat{\mathbf{\Sigma}}_{ww}(m)$, $\hat{\mathbf{V}}$, the standard normal cut-off points and possibly the delta method in approximate inference for functions of the elements of $\mathbf{\Sigma}_{bb}$ and $\mathbf{\Sigma}_{ww}$. As can be seen from Theorems 3, 4, and 6, this procedure is asymptotically justified for cases I and II, if some knowledge on m_0 is available. For case III, the normal distribution based inference is not exactly appropriate. But, for many functions, Wishart based inference is difficult.

By taking into account the higher terms in the common expansion of Theoem 6, our normal approximation is expected to be practically adequate even for relatively small n.

6. SIMULATION

A simulation study was conducted to assess finite sample properties of the asymptotic inference procedures developed in sections 2-5. We considered model (1.1) with p=4 and normally distributed $\mathbf{b}_i$ and $\mathbf{w}_{ij}$. We set $\mu = 0$, but μ was estimated. For the sample configuration (n, r), we chose three sets, $(50, 5)$, $(50, 50)$, and $(5, 50)$ loosely corresponding to cases I, II, and III used to develop asymptotic theory. Note that in many applications, $n = 50$ is not necesarily considered very large and $r = 50$ is unusually large. For the covariance components $\mathbf{\Sigma}_{bb}$ and Σ_{ww}, we considered two parameterizations both of which have rank $\mathbf{\Sigma}_{bb} = m_0 = 2$.

$$\text{(i) } \mathbf{\Sigma}_{bb}^{(\text{i})} = \begin{pmatrix} 1 & 1 & 0.5 & 0.5 \\ 1 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 1 \\ 0.5 & 0.5 & 1 & 1 \end{pmatrix}, \quad \mathbf{\Sigma}_{ww}^{(\text{i})} = \begin{pmatrix} 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{pmatrix},$$

$$\text{(ii) } \mathbf{\Sigma}_{bb}^{(\text{ii})} = \mathbf{\Sigma}_{bb}^{(\text{i})}, \qquad \mathbf{\Sigma}_{ww}^{(\text{ii})} = 30\mathbf{\Sigma}_{ww}^{(\text{i})}.$$

One way to characterize the parameter sets is to consider the roots of $\left|\Sigma_{bb} - \gamma\Sigma_{ww}\right| = 0$. For (i), $\gamma = 1.714,\ 1.333,\ 0,\ 0$ and for (ii), $\gamma = 0.057,\ 0.044,\ 0,\ 0$. Thus, (ii) can be considered to be a case where $\mathbf{\Sigma}_{bb}$ of rank 2 is relatively close to a rank 1 matrix while the rank of $\mathbf{\Sigma}_{bb}$ in (i) can easily be detected to be 2. We first looked at the sample roots $\hat{\lambda}_i$'s as defined in (1.3) to see differences among the sample configurations and parameterizations. Table 6.1 reports the empirical frequency of the number of $\hat{\lambda}_i$'s larger than one. The true rank of Σ_{bb} being 2 implies that we expect exactly 2 roots to be larger than one. We note that this number is the rank of $\hat{\Sigma}_{bb}(4)$, the estimate with no information on the rank. For parameterization (i), all the samples produced at least 2 roots larger than one, but can produce 3 or more roots larger than one quite often regardless of the sample configuration. For parameterization (ii), the number of roots larger than one tends to be smaller than parameterization (i).

Even though the true rank of Σ_{bb} is 2, we considered 4 different situations where a statistician believes the rank is at most m where $m = 1,2,3,4$. The case with $m = 4$ corresponds to that with no information on the rank. As a summary, we report only on inferences for two parametric functions σ_{bb11}, the $(1,1)$ element of Σ_{bb} and $\tau_1 = \sigma_{bb11}/(\sigma_{bb11} + \sigma_{ww11})$, the intra-class correlation for the first variable. The true values are

$$\begin{aligned}(\sigma_{bb11}, \tau_1) &= (1,\ 0.500), \qquad \text{for (i)},\\ &= (1,\ 0.032), \qquad \text{for (ii)}.\end{aligned}$$

Table 6.2 presents the relative bias (bias divided by the true value) of each of the four estimators corresponding to $m = 1,2,3,4$. For the estimator with $m = 1$, the relative bias is large for parameterization (i), and in general the bias does not necessarily decrase with larger n or r. This result is consistent with the fact that the estimator with $m = 1$ is not a consistent estimator in the sense of section 2. The relative biases of the estimators with $m > 1$ are very similar. For these estimators, the bias is not a serious problem except for parameterization (ii) with either small n or small r. Large relative biases for (ii) with small r are due to a combination of the difficult parameter structure and the small true value of a positive parameter. For parameterization (ii), the estimator with $m = 2$ has a smaller bias then that with $m = 3$ or $m = 4$. Table 6.3 gives the mean square errors of the estimators. The estimator with $m = 1$ has a mean square error which is either large or irregular (with relative to n or r). Once again, the estimators with $m = 2, 3, 4$ exhibit similar behavior. For $m = 3$ or $m = 4$, the estimator is different from that with $m = 2$ only if $\hat{\lambda}_3$ (or $\hat{\lambda}_4$) given in Table 1 is larger than one. Even for such a case, $\hat{\lambda}_3$ contributes to the estimator only through $\hat{\lambda}_3 - 1$, i.e., the part of $\hat{\lambda}_3$ larger than 1. Thus, when this part is small relative to $\hat{\lambda}_1 - 1$ and $\hat{\lambda}_2 - 1$, the differences are small. For the difficult case for estimation, i.e., for (ii) with small n or r, the estimator with $m = 2$ seems to have a smaller mean square error than that with $m = 3$ or 4. Hence, the under-specification of the rank seems to lead to a poor performing estimator, while the over-specification does not seriously hamper point estimation possibly except for the difficult cases with small random effect.

To assess the usefulness of the approximate inference procedure suggested in section 5, confidence intervals with nominal 95 % of coverage were computed. These are based on each estimate, the corresponding standard error using $\hat{\mathbf{V}}$ in section 5, the standard normal percentile (1.96), and the delta method for τ_1. Recall that this procedure based on a higher order expansion and normal approximation is asymptotically valid for cases I and II and its use for case III does not have valid justification. For case III, a valid procedure is possible, using Wishart limiting distribution given in Theorem 3. For σ_{bb11}, such a procedure is the chi-square confidence interval with $n - 1$ degrees of freedom. We also computed such an interval using each of the estimators with $m = 1, 2, 3, 4$ for all (n, r) pairs. For each of these different nominal

95 % confidence intervals (8 for σ_{bb11} and 4 for τ_1), the percentage of containing the true value over 1000 replications was obtained. Table 6.4 gives such results. As expected, the intervals based on the estimation under the rank at most 1 have very poor coverage. The differences among those based on $m = 2, 3, 4$ are small for either $\hat{\mathbf{V}}$-based or χ^2_{n-1}-based procedures. Comparing these two approaches with $m = 2, 3, 4$, the $\hat{\mathbf{V}}$-based procedures have larger (sometimes much larger) coverage than the χ^2-based, except for parameterization (i) with $(n, r) = (5, 50)$. The χ^2-based procedure is justifiable for large r cases, but does not seem to perform very well for the difficult case with small random effect (relative to error) and for small r cases. The $\hat{\mathbf{V}}$-based normal interval, originally suggested as a possible procedure regardless of the sampling configuration, in fact provides good coverage properties over different cases, except for parameterization (ii) with small n. For small n, the use of t cut-off points with $n-1$ or $n-1-m$ degrees of freedom would improve the coverage. For parameterization (ii), the over-specified rank (m=4) tends to decrease the coverage. Once the maximum possible rank of a covariance component is reliably established [e.g, using a test procedure in Anderson and Amemiya (1991)], we recommend estimation under the rank condition and approximate inference based on $\hat{\mathbf{V}}$ and normal or t cut-off points.

Table 6.1: Frequency of the number of roots $\hat{\lambda}_i$ greater than one.

	Parameterization					
	(i)			(ii)		
(n, r)	(50, 5)	(50, 50)	(5, 50)	(50,5)	(50, 50)	(5, 50)
number						
0	0	0	0	0	0	1
1	0	0	0	0	60	126
2	212	262	677	599	269	722
3	703	688	321	331	670	151
4	85	70	2	10	61	0

Table 6.2: The relative biases of the estimators with $m = 1, 2, 3, 4$. (bias / true value)

	Parameterization					
	(i)			(ii)		
(n, r)	(50,5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
σ_{bb11} estimator						
m=1	-0.294	-0.292	-0.166	0.184	-0.274	0.007
m=2	0.012	0.006	0.019	0.591	0.022	0.200
m=3	0.019	0.007	0.020	0.637	0.038	0.208
m=4	0.019	0.007	0.020	0.638	0.039	0.208
τ_1 estimator						
m=1	-0.322	-0.317	-0.291	0.171	-0.279	-0.022
m=2	-0.007	-0.007	-0.097	0.577	0.019	0.171
m=3	0.	-0.007	-0.096	0.620	0.034	0.176
m=4	0.	-0.007	-0.096	0.623	0.034	0.176

Table 6.3: The mean square errors of the estimators with $m = 1, 2, 3, 4$. ($\times 100$ for τ_1)

	Parameterization					
	(i)			(ii)		
(n, r)	(50,5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
σ_{bb11} estimator						
m=1	0.20273	0.24504	0.62424	1.45636	0.29814	1.17939
m=2	0.05809	0.04340	0.51607	1.75602	0.10434	1.18803
m=3	0.05788	0.04340	0.51701	1.80604	0.10435	1.18970
m=4	0.05789	0.04340	0.51601	1.80779	0.10436	1.18970
τ_1 estimator						
m=1	5.93730	5.53200	6.94470	0.14366	0.03189	0.10642
m=2	0.45376	0.26559	2.99960	0.17147	0.01007	0.10487
m=3	0.44039	0.26505	2.99310	0.17610	0.00998	0.10494
m=4	0.43971	0.26503	2.99310	0.17622	0.00998	0.10494

Table 6.4: Percentages of the nominal 95 % confidence intervals containing the true values (methods based on $\hat{\mathbf{V}}$ and χ^2_{n-1})

		Parameterization					
		(i)			(ii)		
method	(n, r)	(50, 5)	(50, 50)	(5, 50)	(50, 5)	(50, 50)	(5, 50)
	σ_{bb11} estimator						
	m=1	59.4	54.0	66.2	87.1	64.6	85.1
$\hat{V}$	m=2	93.7	94.1	82.5	95.4	94.2	92.9
	m=3	94.2	94.2	82.5	96.7	95.3	94.5
	m=4	94.3	94.2	82.5	95.2	95.5	89.7
χ^2_{n-1}							
	m=1	54.3	55.9	80.7	19.4	47.1	74.8
	m=2	89.7	94.0	93.8	27.0	79.0	84.1
	m=3	90.0	94.0	93.7	26.9	79.5	84.2
	m=4	89.9	94.0	93.7	26.9	79.5	84.2
	τ_1 interval						
	m=1	53.2	45.3	67.2	85.9	62.4	84.9
$\hat{V}$	m=2	95.1	94.6	86.9	94.6	93.3	92.6
	m=3	94.7	94.6	87.0	95.6	94.6	93.9
	m=4	94.7	94.6	87.0	94.3	94.6	89.2

7. DETAILS

Proof of Theorem 5. We first note that (2.9) in the proof of Theorem 2 and the restriction $m = m_0$ imply

$$P\{\hat{k} = m_0\} \to 1 \tag{7.1}$$

for all three cases. Thus, in deriving an asymptotic expansion, we can consider $\hat{\Sigma}_{bb}(m_0)$ and $\hat{\Sigma}_{ww}(m_0)$ with $\hat{k} = m_0$. It turns out to be easier to derive first the common expansion given in Theorem 6.

Note that for given $\mathbf{C}$ and $\mathbf{D}$ in (4.1), with probability one, $\mathbf{C}\mathbf{b}_i = 0$ for all i and $\mathbf{D}\mathbf{b}_i$ has a nonsingular covariance matrix. Since the result is free of the choice of $\mathbf{C}$ and $\mathbf{D}$, we use, without loss of generality and with possible re-ordering of variables, $\mathbf{C}$ and $\mathbf{D}$ given by

$$\mathbf{C} = (\mathbf{I}_{p-m_0},\ -\beta')', \quad \mathbf{D} = (\mathbf{0},\ \mathbf{I}_{m_0})' - \mathbf{C}\Sigma_{vv}^{-1}\Sigma_{vu},$$

where β is a $m_0 \times (p - m_0)$ nonzero matrix, $\Sigma_{vv} = \mathbf{C}'\Sigma_{ww}\mathbf{C}$, and $\Sigma_{vu} = \mathbf{C}'\Sigma_{ww}(\mathbf{0},\ \mathbf{I}_{m_0})'$. Correspondingly, we assume thart Σ_{bb} has the form

$$\Sigma_{bb} = \begin{pmatrix} \beta \\ \mathbf{I}_{m_0} \end{pmatrix} \Sigma_{xx}(\beta,\ \mathbf{I}_{m_0}),$$

where Σ_{xx} is a $m_0 \times m_0$ symmetric positive definite matrix. Note that $\mathbf{C}'\Sigma_{bb}\mathbf{C} = \mathbf{0}$, rank $\mathbf{D} = m_0$, and $(\mathbf{C},\ \mathbf{D})$ is $p \times p$ nonsingular matrix. Also, $\mathbf{D}\mathbf{b}_i$ has the nonsingular covariance matrix Σ_{xx}.

Using the expansion in Theorem 1 with $\hat{k} = m_0$ and the fact that $\mathbf{m}_{ww} = \Sigma_{ww} + O_p(\frac{1}{\sqrt{nr}})$, we can write

$$\begin{aligned} \hat{\Sigma}_{bb}(m_0) - \Sigma_{bb} &= \frac{1}{r}[\mathbf{m}_{bb} - (r\Sigma_{bb} + \Sigma_{ww})] + O_p(\frac{1}{r\sqrt{n}}), \\ \hat{\Sigma}_{ww}(m_0) - \Sigma_{ww} &= \mathbf{m}_{ww} - \Sigma_{ww} + O_p(\frac{1}{r\sqrt{n}}). \end{aligned} \tag{7.2}$$

Let $\hat{\mathbf{P}} = \mathbf{m}_{ww}^{1/2}\hat{\mathbf{Q}}$, $\hat{\mathbf{T}} = \mathbf{m}_{ww}^{-1/2}\hat{\mathbf{Q}}$, and partition these matrices as

$$\hat{\mathbf{P}} = \begin{pmatrix} \hat{\mathbf{P}}_{11} & \hat{\mathbf{P}}_{12} \\ \hat{\mathbf{P}}_{21} & \hat{\mathbf{P}}_{22} \end{pmatrix}, \qquad \hat{\mathbf{T}} = \begin{pmatrix} \hat{\mathbf{T}}_{11} & \hat{\mathbf{T}}_{12} \\ \hat{\mathbf{T}}_{21} & \hat{\mathbf{T}}_{22} \end{pmatrix},$$

where $\hat{\mathbf{P}}_{11}$ and $\hat{\mathbf{T}}_{11}$ are $(p - m_0) \times m_0$ and $\hat{\mathbf{Q}}$ is defined in (3). Define

$$\begin{aligned} \hat{\Sigma}_{xx} &= \frac{1}{r}\hat{\mathbf{P}}_{21}(\hat{\Lambda}_1 - \mathbf{I}_{m_0})\hat{\mathbf{P}}_{21}', \\ \hat{\beta}' &= (\hat{\mathbf{P}}_{11}\hat{\mathbf{P}}_{21}^{-1})' = -\hat{\mathbf{T}}_{22}\hat{\mathbf{T}}_{12}^{-1}. \end{aligned}$$

Furthermore, if we let $\hat{\mathbf{C}} = (\mathbf{I}_{p-m_0},\ -\hat{\beta}')'$ and $\hat{\mathbf{A}} = \hat{\mathbf{T}}_{12}\hat{\mathbf{T}}_{12}'$, then

$$\begin{aligned} \hat{\mathbf{A}} &= (\hat{\mathbf{C}}'\mathbf{m}_{ww}\hat{\mathbf{C}})^{-1} \\ \Omega_{m_0} &= \mathbf{m}_{ww}\hat{\mathbf{C}}\hat{\mathbf{A}}\hat{\mathbf{C}}'(\mathbf{m}_{bb} - \mathbf{m}_{ww})\hat{\mathbf{C}}\hat{\mathbf{A}}\hat{\mathbf{C}}'\mathbf{m}_{ww}, \end{aligned} \tag{7.3}$$

where Ω_{m_0} is defined in (1.6) with $\hat{k} = m_0$. See Amemiya and Fuller (1984, pp. 449). Let $\mathbf{S}_{xx}$, $\mathbf{S}_{vx}$, $\mathbf{S}_{vv}$, and $\mathbf{S}_{vv.x}$ be as defined in (4.1) with this particular choice of $\mathbf{C}$ and $\mathbf{D}$.

Multiplying the first equation in (7.2) by $(\mathbf{0},\ \mathbf{I}_{m_0})$ on the left and $(\mathbf{0},\ \mathbf{I}_{m_0})'$ on the right, we get for all three cases

$$\hat{\Sigma}_{xx} = \mathbf{S}_{xx} + O_p(\frac{1}{\sqrt{nr}}). \tag{7.4}$$

Multiplying the first equation in (7.2) by $\mathbf{C}'$ on the left and $(\mathbf{0},\ \mathbf{I}_{m_0})'$ on the right, we get for all three cases

$$\hat{\beta} - \beta = \frac{1}{\sqrt{r}}\mathbf{S}_{xx}^{-1}\mathbf{S}_{vx}' + Op(\frac{1}{r\sqrt{n}}). \tag{7.5}$$

Since $\frac{1}{\sqrt{r}}\mathbf{S}_{vx} = Op(\frac{1}{\sqrt{nr}})$, (7.5) implies that

$$\hat{\mathbf{C}} - \mathbf{C} = Op(\frac{1}{\sqrt{nr}}). \tag{7.6}$$

If we write $\hat{\mathbf{C}} = \mathbf{C} - \mathbf{E}$ with $\mathbf{E} = [\mathbf{0},\ (\hat{\beta} - \beta)']'$, then, for all cases

$$\hat{\mathbf{C}}'(\mathbf{m}_{bb} - \mathbf{m}_{ww})\hat{\mathbf{C}} = \mathbf{S}_{vv.x} - \mathbf{S}_{ee} + Op(\frac{1}{n\sqrt{r}}). \tag{7.7}$$

Also, from (7.3), (7.6), and the fact that $\mathbf{m}_{ww} = \mathbf{\Sigma}_{ww} + Op(\frac{1}{\sqrt{nr}})$, we have for all three cases

$$\hat{\mathbf{A}} = \mathbf{\Sigma}_{vv}^{-1} + Op(\frac{1}{\sqrt{nr}}). \tag{7.8}$$

Now, using the second equation in (7.3), (7.6), (7.7), (7.8), and the fact that $\mathbf{m}_{ww} = \mathbf{\Sigma}_{ww} + Op(\frac{1}{\sqrt{nr}})$, it follows that

$$\mathbf{\Omega}_{m_0} = \mathbf{\Sigma}_{ww}\mathbf{C}\mathbf{\Sigma}_{vv}^{-1}(\mathbf{S}_{vv.x} - \mathbf{S}_{ee})\mathbf{\Sigma}_{vv}^{-1}\mathbf{C}'\mathbf{\Sigma}_{ww} + Op(\frac{1}{n\sqrt{r}}). \tag{7.9}$$

The result for case II follows from (7.6) and (7.9). The approximations for cases I and III follow from the fact that $\mathbf{\Omega}_{\mathrm{II}} = \mathbf{\Omega}_{\mathrm{I}} + Op(\frac{1}{n})$ for case I and $\frac{1}{r}(\mathbf{m}_{ww} - \mathbf{\Sigma}_{ww}) = Op(\frac{1}{r\sqrt{r}})$ for case III.

8. CONCLUSION

For the multivariate balanced random effect model, estimation for the covariance components was discussed . The estimators considered are the restricted (residual) maximum likelihood estimators derived under the assumption that the rank of the between-group covariance matrix is at most a specified number. By deriving the asymptotic properties under a possibly incorrectly specified rank, we found that the true unknown rank has a large influence in determining the behavior of the estimators. Based on the higher order expansion, an approximate inference procedure useful for various situations was developed. The simulation study supported the usefulness of the procedure.

REFERENCES CITED

Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite. Amer. Statist. 39, 112-117.

Amemiya, Y., and W. A. Fuller (1984). Estimation for the multivariate Errors-in-variables model with estimated error covariance matrix. Ann. Statist. 12, 497-509.

Amemiya, Y., T. W. Anderson, and P. A. W. Lewis (1990). Percentage points for a test of rank in multivariate components of variance. Biometrika 77, 637-641.

Anderson, B., T. W. Anderson, and I. Olkin (1986). Maximum likelihood estimators and likelihhod ratio criterion in multivariate components of variance. Ann. Statist. 14, 405-417.

Anderson, T. W. (1984). Estimating linear statistical relationships. Ann. Statist. 12, 1-45.

Anderson, T. W. (1989). The asymptotic distribution of the likelihood ratio criterion for testing the rank in multivariate components of variance. J. Multivariate Anal. 30, 72-79.

Anderson, T. W., and Y. Amemiya (1991). Testing dimensionality in the multivariate analysis of variance. Statist. Probab. Lett. 12, 445-463.

Bhargava, A. K., and D. Disch (1982). Exact probabilities of obtaining estmated nonpositive definite between-group covariance matrices. J. Statist. Computing and Simulation, 15, 27-32.

Bock, R. D., and A. C. Petersen (1975). A multivariate correction for attenuation. Biometrika, 62, 673-678.

Calvin, J. A., and R. L. Dykstra (1991). Maximum likelihood estimation of a set of covariance matrices under lowner order restrictions with applications to balanced multivariate variance components models. Ann. Statist. 19, 850-869.

Feller, W. (1970). An introduction to probability ant its applications. Vol. II, 2nd ed. Wiley, New York.

Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems. J. Amer. Statist. Assoc. 72, 320-338.

Hill, W. G., and R. Thompson (1978). Probabilities of nonpositive definite between-group or genetic covariance matrices. Biometrics 34, 429-439.

Klotz, J. H., and J. Putter (1969). Maximum likelihood estimation of multivariate covariance components for the balanced one-way lay-out. Ann. Math. Statist. 40, 1100-1105.

Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics 41, 153-165.

Miller, J. J. (1977). Asymptotic properties of maximum likelihhood estimates in the mixed model of the analysis of variance. Ann. Statist. 5, 746-762.

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. Ann. Statist. 1, 763-765.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. Statist. Science 6, 15-51.

Schott, J. R., and J. G. Saw (1984). A multivariate one-way classification model with random effects. J. multivariate Anal. 15, 1-12.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). Variance components. Wiley, New York.

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. Biometrics, 22, 527-550.

OVERALLL SUMMARY AND CONCLUSION

This dissertation consisting of three papers discussed problems arising in, or related to, the statistical ananlysis based on the multivariate mixed effect model. Throughtout, the detection and implication of a possibly reduced rank structure for an underlying random effect were the major topics.

In the first paper, the limiting distribution of the roots of a certain determinantal equation was derived under the general assumption of differential convergence rates. Although the assumption includes that corresponding to the multivariate mixed effect model. the result was derived and presented in a general form. The connection to the covariance component problem was illustrated by considering application to a simple model.

The second paper dealt with the problem of testing for the rank of a covariance component in the general multivariate mixed effect model. Asymptotic properties of the propsoed test procedures were discussed under various assumptions. For a simple case, the exact null distribution was characterized. Test procedures which can be useful for practical situations were proposed. The simulation study showed the usefulness of the procedures.

In the third paper, estimation of the covariance components in the multivariate one-way random effect model was discussed. The behavior of the restricted (residual) maximum likelihood estimator was shown to be heavily dependent on the unknown true rank of the between-group covariance component. An approximate inference procedure based on the estimator derived under a rank constraint and on the higher order expansion was proposed as a means to deal with various practical situations. The practicality of the procedure was also shown by the simulation study.

REFERENCES

Amemiya, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite. Amer. Statist. 39, 112-117.

Anderson, T. W., and Y. Amemiya (1991). Testing dimensionality in the multivariate analysis of variance. Statist. Probab. Lett. 12, 445-463.

Greene, W. H. (1990). Econometric Analysis. MacMillan Publishing Company. New York.

Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and to related problems. J. Amer. Statist. Assoc. 72, 320-338.

Henderson, C. R. (1953). Estimation of variance and covariance components. Biometrics 9, 226-252.

Herbach, L. H. (1959). Properties of model II type analysis of variance tests, A: optimum nature of the F-test for model II in the balanced case. Ann. Math. Statist. 30, 939-959.

Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. Biometrics 41, 153-165.

Miller, J. J. (1977). Asymptotic properties of maximum liklihood estimates in the mixed model of the analysis of variance. Ann. Statist. 5, 746-762.

Patterson, H. D., and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

Searle, S. R., G. Casella, and C. E. McCulloh (1992). Variance components. Wiley, New York.

Thompson, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. Biometrics 22, 527-550.

Thompson, W. A. Jr. (1962). The problem of negative estimates of variance components. Ann. Math. Statist. 33, 273-289.